(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/31117 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/32202

(22) International Filing Date: 15 October 2001 (15.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/687,837	13 October 2000 (13.10.2000)	US

(71) Applicants *(for all designated States except US)*: **ARBOR VITA CORPORATION** [US/US]; 772 Lucerne Drive, Sunnyvale, CA 94085 (US). **GARMAN, Jonathan, David** [US/US]; 5084 McCoy Avenue, San Jose, CA 95130 (US). **CANDIA, III, Albert, Frederick** [US/US]; 789 Marsh Road, Menlo Park, CA 94025 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **LU, Peter, S.** [US/US]; 99 East Middlefield Road, No. 29, Mountain View, CA 94043 (US).

(74) Agents: **SERAFINI, Andrew, T.** et al.; Townsend and Townsend and Crew LLP, 2 Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

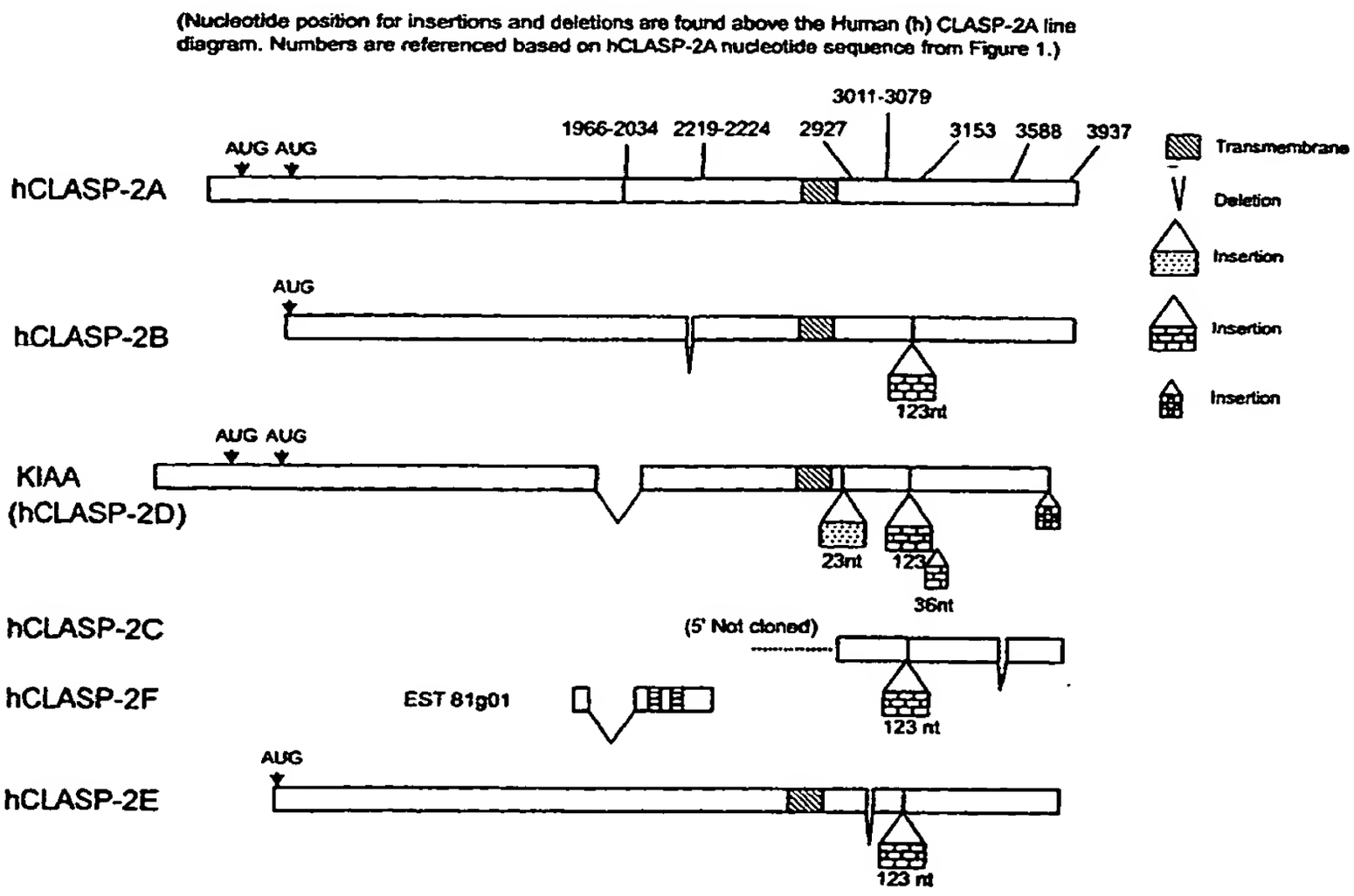
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: CLASP-2 TRANSMEMBRANE PROTEINS



(57) Abstract: The present invention relates to a cell surface molecule, designated cadherin-like asymmetry protein-2 ("CLASP-2"). In particular, it relates to CLASP-2 polynucleotides, polypeptides, fusion proteins, and antibodies. The invention also relates to methods of modulating an immune response by interfering with CLASP-2 function.

WO 02/31117 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CLASP-2 TRANSMEMBRANE PROTEINS

0.0 CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part application of U.S. Patent Application No. 09/547,276 filed April 11, 2000, which claims the benefit of U.S. Provisional Application Nos. 60/182,296 filed February 14, 2000, 60/176,195 filed January 14, 2000, 60/170,453 filed December 13, 1999, 60/162,498 filed October 29, 1999, 60/160,860 filed October 21, 1999, 60/134,118 filed May 14, 1999, 60/134,117 filed May 14, 1999, 60/134,114 filed May 14, 1999, and 60/129,171 filed April 14, 1999, the disclosures of which are incorporated by reference.

1.0 FIELD OF THE INVENTION

The present invention relates to molecules expressed in cells of the immune system. In particular, the invention relates to a transmembrane protein that contains certain classical cadherin characteristics.

2.0 BACKGROUND OF THE INVENTION

The generation of an immune response against an antigen is carried out by a number of distinct immune cell types that work in concert within the context of a particular antigen. The helper T cell (T_H) plays a pivotal role to coordinate two types of antigen-specific immune response; *i.e.*, cellular and humoral immune response. Recognition of antigen by T cells requires the formation of a specialized junction between the T cell and the antigen-presenting cell (APC) called the "immulogical synapse" (Dustin, *et al.*, 1998, Cell 94: 667-677). The immune synapse orchestrates recruitment and exclusion of specific proteins from the contact area by an unknown mechanism and is thought to be initiated by T-cell antigen receptor (TCR) recognition of peptides bound to MHC molecules (antigen) (Monk, *et al.* 1998, Nature 395: 82). However, the low affinity of the TCR for antigen as well as limited number of ligands makes it unlikely that TCR: antigen interaction alone is sufficient to drive the formation of the immunological synapse (Matsui *et al.*, 1994, Proc. Natl. Acad. Sci. U.S.A. 91: 12861-12866).

Costimulatory molecules such as CD4, ICAM-1, LFA-1, CD28, CD2 have been proposed to stabilize the cell-cell contact (Dustin, *et al.*, 1999, Science 283: 649). However, since these molecules are recruited to the synapse after activation they cannot

account for the high specificity and avidity during the early phases of T-cell antigen recognition. Recent work demonstrated that a portion of the T cell surface at the leading edge is specialized to mediate the early phases of synapse formation (Negulescu, *et al.*, 1996, Immunity 4: 421-430). Such a specialization must be a pre-formed structure containing cell surface adhesion proteins (ectodomains) to augment TCR engagement and corresponding cytoplasmic portions (endodomains) to transduce signals and bind cytoskeleton to maintain structural/functional polarity.

The ectodomain of the pre-formed synapse or "immune gateway" was recently discovered and is created in part by CLASP-1 (U.S.S.N. 09/411,328, filed October 1, 1999; PCT/US99/22996). In addition to cadherin motifs, CLASP-1 also contains a CRK-SH3 binding domain, tyrosine phosphorylation sites, and coiled/coil domains suggesting direct interaction with cytoskeleton and regulation by adaptor molecules such as CRK. The *CLASP-1* transcript is present in lymphoid organs and neural tissue, and the protein is expressed by T and B lymphocytes and macrophages in the MOMA-1 subregion of the marginal zone of the spleen, an area known to be important in T: B cell interaction. CLASP-1 staining of individual T and B cells exhibits a preactivation structural polarity, being organized as a "ball" or "cap" structure in B cells, and forming a "ring", "ball" or "cap" structure in T cells. The placement of these structures is adjacent to the microtubule-organizing center ("MTOC"). CLASP-1 antibody staining indicates that CLASP-1 is at the interface of T-B cell conjugates that are fully committed to differentiation. Antibodies to the extracellular domain of CLASP-1 also block T-B cell conjugate formation and T cell activation.

3.0 SUMMARY OF THE INVENTION

The present invention relates to a cell surface molecule, a member of a new multigene family designated cadherin-like asymmetry protein(s) ("CLASP(s)"). In particular, it relates to a polynucleotide comprising a coding sequence for CLASP-2, a polynucleotide that selectively hybridizes to the complement of a CLASP-2 coding sequence, expression vectors containing such polynucleotides, genetically-engineered host cells containing such polynucleotides, CLASP-2 polypeptides, CLASP-2 fusion proteins, therapeutic compositions, CLASP-2 domain mutants, antibodies specific for CLASP-2 polypeptides, methods for detecting the expression of CLASP-2, and methods of inhibiting an immune response by interfering with CLASP-2 function. A wide variety of uses are encompassed by the invention, including but not limited to, treatment of autoimmune

diseases and hypersensitivities, prevention of transplantation rejection responses, and augmentation of immune responsiveness in immunodeficiency states.

In one aspect, the invention provides an isolated CLASP-2 polynucleotide that is: (a) a polynucleotide that has the sequence of SEQ ID NO: 1, 3, 5 or 9; (b) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and encodes a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10 or an allelic variant or homologue of a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10; or (c) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and encodes a polypeptide with at least 25 contiguous residues of the polypeptide of SEQ ID NO: 2, 4, 6 or 10; or (d) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and has at least 12 contiguous bases identical to or exactly complementary to SEQ ID NO: 1, 3, 5, or 9. In a related aspect, the invention provides a CLASP-2 polynucleotide wherein the polynucleotide encodes a polypeptide that binds to the PDZ domain of PSD95, DLG1 or neDLG. 2. In another related aspect, the invention provides a CLASP-2 polynucleotide wherein the polynucleotide encodes a polypeptide that has a binding affinity of at least 10^4 M^{-1} for binding PSD95, DLG1 or neDLG.

In one aspect, the invention provides a CLASP-2 polynucleotide that encodes a polypeptide having the full-length sequence of SEQ ID NO: 2, 4, 6, or 10 or the cDNA sequence encoded by the inserts of ATCC Deposit Nos: PTA-1562, PTA-1563 and PTA-1573.

In another aspect, the invention provides a CLASP-2 polynucleotide that encodes a polypeptide having the full-length sequence of Isoform 1, Isoform 2, or Isoform 3 (SEQ ID NO: _____) or the cDNA sequence encoded by the inserts of AVC-PD14 (ATCC Deposit No. _____) and AVC-PD19 (ATCC Deposit No. _____).

In another aspect, the invention further provides an isolated CLASP-2 polynucleotide comprising a nucleotide sequence that has at least 90% percent identity to SEQ ID NO: 1, 3, 5 or 9 as calculated using FASTA wherein said sequences are aligned so that highest order match between said sequences is obtained.

The invention further provides an isolated polypeptide comprising a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 2, 4, 6 or 10 and is immunologically crossreactive with SEQ ID NO: 2, 4, 6 or 10 or shares a biological function with native CLASP-2.

The invention also provides vectors, such as expression vectors, comprising a polynucleotide sequence of the invention In other embodiments, the invention provides host

cells or progeny of the host cells comprising a vector of the invention. In certain embodiments, the host cell is a eukaryote. In other embodiments, the expression vector comprises a CLASP-2 polynucleotide in which the nucleotide sequence of the polynucleotide is operatively linked with a regulatory sequence that controls expression of the polynucleotide in a host cell. In certain embodiments, the invention provides a host cell comprising a CLASP-2 polynucleotide, wherein the nucleotide sequence of the polynucleotide is operatively linked with a regulatory sequence that controls expression of the polynucleotide in a host cell, or progeny of the cell.

In another aspect, the invention further provides a CLASP-2 polynucleotide that is an antisense polynucleotide. In a preferred embodiment, the antisense polynucleotide is less than about 200 bases in length. In other embodiments, the invention provides an antisense oligonucleotide complementary to a messenger RNA comprising SEQ ID NO: 1, 3, 5 or 9 and encoding CLASP-2, wherein the oligonucleotide inhibits the expression of CLASP-2.

In another aspect, the invention provides an isolated DNA that encodes a CLASP-2 protein as shown in SEQ ID NO: 2, 4, 6 or 10. In certain embodiments, the CLASP-2 polynucleotide is RNA.

The invention provides a method for producing a polypeptide comprising: (a) culturing the host cell containing a CLASP-2 polynucleotide under conditions such that the polypeptide is expressed; and (b) recovering the polypeptide from the cultured host cell or its cultured medium.

The invention further provides an isolated CLASP-2 polypeptide encoded by a CLASP-2 polynucleotide. In some embodiments, the CLASP-2 polypeptide has the amino acid sequence of SEQ ID NO: 2, 4, 6 or 10, or a fragment thereof. In some embodiments, the isolated CLASP-2 polypeptide is cell-membrane associated. In other embodiments, the isolated CLASP –2 polypeptide is soluble. In other embodiments, the soluble CLASP-2 polypeptide is fused with a heterologous polypeptide.

The invention further provides an isolated CLASP-2 protein having the sequence as shown in SEQ ID NO: 2, 4, 6 or 10. In some embodiments, the invention provides a CLASP-2 protein comprising the sequence as shown in SEQ. ID. NO: 1 and variants thereof that are at least 95% identical to SEQ ID. NO: 2 and specifically binds a cytoskeletal protein. In certain embodiments the cytoskeletal protein is spectrin.

The invention further provides an isolated antibody that specifically binds to a polypeptide having the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 10, or a

binding fragment thereof. In some embodiments the antibody is monoclonal. In other embodiments, the invention provides a hybridoma capable of secreting the antibody.

The invention further provides a method of identifying a compound or agent that binds a CLASP-2 polypeptide comprising: i) contacting a CLASP-2 polypeptide with the compound or agent under conditions which allow binding of the compound to the CLASP-2 polypeptide to form a complex and ii) detecting the presence of the complex.

The invention further provides a method of detecting a CLASP-2 polypeptide in a sample, comprising: (a) contacting the sample with a CLASP-2 antibody or binding fragment and (b) determining whether a complex has been formed between the antibody and with CLASP-2 polypeptide.

The invention further provides a method of detecting a CLASP-2 polypeptide in a sample, comprising: (a) contacting the sample with a CLASP-2 polynucleotide or a polynucleotide that comprises a sequence of at least 12 nucleotides and is complementary to a contiguous sequence of the CLASP-2 polynucleotide and (b) determining whether a hybridization complex has been formed.

The invention further provides a method of detecting a CLASP-2 nucleotide in a sample, comprising: (a) using a polynucleotide that comprises a sequence of at least 12 nucleotides and is complementary to a contiguous sequence of a CLASP-2 polynucleotide in an amplification process; and (b) determining whether a specific amplification product has been formed.

The invention further provides pharmaceutical compositions comprising a CLASP-2 polynucleotide, a CLASP-2 polypeptide, or a CLASP-2 antibody and a pharmaceutically acceptable carrier.

In one aspect, the invention provides a method of inhibiting an immune response in a cell comprising: (a) interfering with the expression of a CLASP-2 gene in the cell; (b) interfering with the ability of a CLASP-2 protein to mediate cell-cell interaction (*e.g.*, interfering with a heterotypic and/or homotypic interaction) between CLASP-2 and an extracellular protein; (c) interfering with the ability of a CLASP-2 protein to bind to another protein. In some such methods, the cell is a T cell or a B cell. Some such methods comprise contacting the cell with an effective amount of a polypeptide which comprises the amino acid sequence of SEQ ID NO: 2, 4, 6 or 10 or a fragment thereof.

In another aspect, the invention provides a method of inhibiting an immune response in a subject, comprising administering to the subject a therapeutically effective

amount of an antibody which specifically binds a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10.

In another aspect, the invention provides a method of preventing or treating a CLASP-2-mediated disease comprising administering to a subject in need thereof a therapeutically effective amount of a CLASP-2 pharmaceutical composition. In some such methods, the CLASP-2-mediated disease is an autoimmune disease.

The invention further provides a method of treating an autoimmune disease in a subject caused or exacerbated by increased activity of T_H1 cells consisting of administering a therapeutically effective amount of a CLASP-2 pharmaceutical composition to the subject.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Nucleotide and predicted amino acid sequence of CLASP-2A cDNA. Notable protein motifs are indicated above the nucleotide sequence in bold. Potential initiator methionines are underscored. The notable, predicted protein motifs are: a cadherin cleavage site encoded by nucleotides 854-868, a cadherin ectodomain (EC) encoded by nucleotides 1253-1264, a transmembrane domain encoded by nucleotides 2861-2917, a coiled coil domain encoded by nucleotides 3579-3682, a second coiled coil domain encoded by nucleotides 3827-3937, and a PDZ binding motif (PBM) encoded by nucleotides 4046-4057.

Figure 2. A. Schematic of CLASP-2 splice variants. Splice variants are compared to Human (h) CLASP-2A. Numbers above hCLASP-2A line diagram indicate where splice variations comprising deletions and insertions relative to hCLASP-2A are found. Abbreviations: "KIAA" KIAA1058 sequence (Genbank Accession No. AB028981). **B.** Nucleotide and predicted amino acid sequence of CLASP-2A cDNA. Notable protein motifs are indicated above the nucleotide sequence in bold. Exact position of insertions and deletions are indicated above the CLASP-2A sequence with arrows and "x", respectively. The nucleotide sequence of insertions schematized in FIG. 2A are indicated above the arrow. The insertions and deletions are as follows (numeration refers to Human CLASP-2A nucleotide sequence): Nucleotides 1966-2034 are deleted in CLASP-2D. Nucleotides 2219-2224 are deleted in CLASP-2B. There is an insertion of 69 amino acids at nucleotide 2927 found in CLASP-2D. The nucleotide sequence for this insertion is: AAGCAGTCCAGTGGGAGCCGCCCCTTCTCCCCCACAGCCATAGCGCCTGCCTGAG GAGGAGCCGGGGAG and encodes amino acids AVQWEPPLLPHSHSACLRRSRG (one letter amino acid abbreviation). This amino acid sequence encodes a putative SH3 binding

domain. There is another deletion at between nucleotides 3011-3079 found in CLASP-2E. CLASPs 2B, 2C, 2D and 2E contain an insertion at nucleotide 3153 with the nucleotide sequence of:

TGAGAGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGAC
CGAGGTCATGCACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTC
TTCGGGCAGGCAGCGCAATACCAGTTTACAGACAGTGAAACAGATGTGGAGGGA
TT. The entire sequence is found in CLASP-2D and encodes amino acids
ERLAHLYDTLHRAYSKVTEVMHSGRRLLGTYFRVAFFGQAAQYQFTDSETDVEG
while the underline sequence is found in CLASPs 2B, 2C, and 2E and encodes amino acids
ERLAHLYDTLHRAYSKVTEVMHSGRRLLGTYFRVAFFGQG. This amino acid sequence encodes a putative immunoreceptor tyrosine-based activation motif (ITAM). There is a two nucleotide deletion in Human CLASP-2C found at nucleotides 3586 and 3587. There is an insertion of 8 nucleotides found only in Human CLASP-2D with sequence: CTGGGATG at nucleotide 3937. This insertion puts a stop codon into the CLASP-2D nucleotide sequence.

Figure 3. A. Alignment of nucleotide sequences of the CLASP-2 isoforms. Sequences were aligned using ClustalW **B**. Alignment of amino acid sequence of the CLASP-2 isoforms. Sequences were aligned using ClustalW. One letter amino acid abbreviation is used.

Figure 4. Expression of CLASP-2 in human cell lines and human tissues as determined by Northern hybridization. A CLASP-2-specific DNA fragment was generated by PCR from a CLASP-2 cDNA clone (HC2-5'), using primers HC2AS2 and HC2S1. The fragment was labeled by incorporation of radioactive ^{32}P dCTP. **A.** Expression in human tissues. The labeled DNA fragment was used as a probe on a human Multiple Tissue Northern (Clonetech MTN Blot, #7780-1). A single band is clearly detect migrating at approximately 7.5 kb in placenta, heart kidney and lung in the Multiple Tissue Northern. Slight expression is detected in liver, skeletal muscle and brain. **B.** Expression in hematopoietic cell lines. A Northern with RNA from multiple cells lines was hybridized with the same hCLASP-2 probe. A similarly migrating band is detected in Jurkat (T-cell derived), 9D10 (B-cell derived) and 293 (human kidney derived) cell lines. There are multiple weaker bands in the 9D10 lane indicating possible splice variants of hCLASP-2. Weak expression is also detected in the mouse cell lines CH27 (B cell lymphoma) and 3A9 (T-cell hybridoma).

Since hybridization and washing were carried out at high stringency, this indicates that the human CLASP-2 probe may cross-react with mouse CLASP mRNA.

Figure 5. A. Amino acid sequence of human and rat CLASP proteins. Sequences were aligned using ClustalW. One letter amino acid abbreviation used. Protein motifs are found within the labeled boxes. A "-" indicates gaps that are placed to acquire a best overall alignment. Other abbreviations: "HC2A" Human CLASP-2 sequence, "KIAA" KIAA1058 sequence (Genbank Accession No. AB028981), "rat" TRG gene (Genbank Accession No. X68101), "HC4" Human CLASP-4 sequence, "HC1" Human CLASP-1 sequence, "HC3" Human CLASP-3 sequence, "HC5" Human CLASP-5 sequence. **B.** Alignment of DOCK motifs found within the human CLASPs and compared to canonical DOCK motifs. Consensus amino acids found within all DOCK motifs are also indicated.

Figure 6. A. Nucleotide and predicted amino acid sequence of CLASP-2A cDNA. Notable protein motifs are indicated (see FIG. 1 legend for details). Additionally, boundaries between exons and introns are indicated by arrows. These boundaries were defined by sequencing Bacterial Artificial Chromosomes (BACs) containing genomic DNA corresponding to CLASP-2. BACs were sequenced using primers derived from exon sequences corresponding to the CLASP-2 cDNA. Each exon/intron boundary is noted (as "Ref" with an appropriate reference number) above the cDNA sequence. The references contain exact nucleotide location of introns. The names and nucleotide numbers of the primers that were used in sequence reactions are also indicated. All nucleotide numbers refer to CLASP-2A cDNA sequence. As shown in the reference, not all of the sequence from sequencing reactions produced sequence matching the cDNA. These nucleotide sequences that did not match the exon sequence for CLASP-2 were considered to be intron sequences. **B.** Alignment of human and rat CLASP amino acid sequences by ClustalW. Notable protein motifs are indicated (see FIG. 1 Legend for additional details). Additionally, the exon/intron borders described in part A are indicated with vertical lines between appropriate amino acids. Reference numbers are indicted in the right margin and correspond to references in Fig 6A and B.

Figure 7. Southern hybridization analysis of CLASP-2. Genomic DNA from HeLa cells or a BAC DNA clone was digested with EcoRI or HinDIII (genomic DNA) or Pst I (BAC DNA) and eletrophoresed and transferred to nylon membrane by standard methods. For a probe, a CLASP-2-specific DNA fragment was generated by PCR from a CLASP-2

cDNA clone (HC2-5'), using primers HC2AS2 and HC2S1. The fragment was labeled by incorporation of radioactive ^{32}P dCTP. Probe HC2.1 is 800 bp long and it recognizes two fragments (~4.5 kb and 1.85 kb) on Eco RI digested genomic DNA. Three fragments are revealed by this probe when hybridized to digested DNA of BACs 4 and 6, with the two major ones identical in size to those detected on genomic DNA.

Figure 8. Expression of human CLASP-1 (hCLASP-1) CLASP-1 and CLASP-2 Glutathion-S-Transferase (GST) fusion proteins. Nucleotides encoding a portion of the hCLASP-2A intracellular domain (nucleotides 3230-4065) were subcloned into pGEX vectors (Pharmacia). Recombinant plasmids were transformed into E. coli (strain DH5α), and transformed strains were grown by standard conditions. While in log phase cells were either induced (I) with IPTG (0.1 mM final concentration) or left uninduced (U). After several additional hours of growth cells were harvested and soluble protein lysates generated by standard methods. Aliquots of the protein lysates were eletrophoresed on SDS-PAGE along with molecular mass standards. The gel was stained with Coomassie Blue and shows that fusion proteins migrated with their predicted molecular masses of 59 and 57 kD for hCLASP-1 and hCLASP-2, respectively.

Figure 9. A. Binding of CLASP-2 C-terminal 20 amino acids to PDZ domains. 20 μM biotinylated synthetic peptide corresponding to the C-terminal 20 amino acids of CLASP-2 was reacted with the indicated plate bound GST fusion proteins (none = no GST fusion protein coated onto plate). Error bars indicate standard deviation of duplicate measurements. **B.** Affinity of CLASP 2 – PDZ interactions. Varying concentrations of biotinylated CLASP-2 peptide were reacted with plate bound GST alone, GST-DLG1, GST-NeDLG, and GST-PSD95 fusion proteins. The binding to GST alone (< 0.1 OD units) was subtracted from the binding to the fusion proteins and the remaining signal was divided by the signal observed upon addition of 30 μM CLASP-2 peptide to each PDZ domain-containing protein (0.4 – 1.0 OD units) and plotted. The plotted data was fit to a saturation binding curve, yielding an apparent affinity of 7.5 μM for NeDLG- CLASP-2 interaction, 21 μM for DLG1- CLASP-2 interaction, and 45 μM for PSD95-CLASP-2 interaction. Data are means of duplicate data points, with standard errors between duplicate data points < 10%. **C.** Inhibition of CLASP-2 – PDZ binding. 5 μM biotinylated synthetic peptide corresponding to the C-terminal 20 amino acids of CLASP-2 was reacted with the indicated, plate-bound PDZ domain-containing GST fusion proteins in the presence or absence of 100 μM competitor

peptide. CLASP-2 Inhibitor refers to a synthetic peptide composed of the eight C-terminal amino acids of CLASP-2. KV1.3 Inhibitor refers to a synthetic peptide composed of the 19 C-terminal amino acids of KV1.3, a lymphocyte potassium channel. The amino acid sequence of the KV1.3 inhibitor is TTNNNPNSAVNIKKIFTDV. **D**. Inhibition of KV1.3 – PDZ binding. 5 μM biotinylated synthetic peptide corresponding to the C-terminal 19 amino acids of KV1.3 was reacted with the indicated, plate-bound PDZ-domain containing GST fusion proteins in the presence or absence of 100 μM CLASP-2 Inhibitor (see FIG. 9C legend).

Figure 10. Preliminary nucleotide sequences of CLASP-2 cDNAs.

Figure 11. A) Full length cDNA sequence and predicted amino acid translation of the human CLASP-2 gene. Predicted initiator methionine starts at nucleotide +1. Three independent 1st exons (indicated as 11A, 11B and 11C) splice into the second exon starting at nucleotide -101. The sequence appearing in FIG. 1 corresponds to nucleotides 1884 through 6690 of FIG 11A. **B)** Differences between the human CLASP-2 cDNA isoforms. In addition to the differential first exon usage indicated in **A**, sequencing multiple, independent cDNA products revealed nucleotide polymorphisms (allelic variations) between CLASP-2 cDNA isoforms. Additionally, differential exon usage through alternative splicing events was discovered. The use of the exon in **B** leads to a premature stop codon that can generate a soluble form of CLASP-2. **C**. Schematic of human CLASP-2 cDNAs. The top line represents nucleotide numbering found in FIG. 11A. Line **(i)** represents CLASP-2 cDNA shown in FIG. 1 above; line **(ii)** represents the full length CLASP-2 isoforms, where there are three CLASP-2 full length cDNA isoforms (A + Z, B + Z, and C + Z). Each of the isoforms uses a unique first exon (A, B, and C) (*see* FIG. 11A) that splices into the rest of the cDNA from exon 2 onwards represented by Z. The portion of the cDNA represented by Z itself has alternative splice and nucleotide polymorphisms that are shown in FIG. 2 above. Line **(iii)** represents the additional 5' sequence with a small region of overlao between nucleotides 1884 to 2109 in FIG. 11A and nucleotides 1-225 of FIG. 1.

Figure 12. Sequence of human CLASP-2 exons and intron boundaries. **A** Sequence of human CLASP-2 exons and intron borders. Stretches of noncontigous genomic sequence from the Human Genome Project (GENBANK entry gi9988160) were aligned using the human CLASP-2 cDNA as a template and Sequencher sequence analysis software (Gene Codes Corp). 22 exons representing approximately the 5' 20% of the human CLASP-

2 cDNA sequence are presented in predicted 5' to 3' order. Exon sequences are underlined and are flanked by intron sequence. Nucleotide numbers in parentheses refer to the exon sequence within the uniquely-generated, contiguous gi9988160 sequence, which is located in B. B. Ordered stretch of human genomic DNA at the CLASP-2 locus aligned from noncontiguous, shotgun sequencing from the Human Genome Project using the human CLASP-2 sequence from FIG. 5A to determine genomic DNA fragment order and orientation.

Figure 13. Amino acid alignment and comparison between the human (h) CLASP family members. Amino acid sequences were aligned using ClustalW. The alignment is presented in order of their greatest pairwise similarity scores. Single letter amino acid abbreviations are used. Astericks indicate complete identity, while colons and periods indicate sequence similarity among CLASP family members. Dashes indicate gaps inserted in the amino acid sequence to facilitate alignment. Labelled boxes are domains with similarity to known protein motifs; unlabelled boxes represent regions of similarity between all CLASPs and may represent CLASP-specific domains.

Figure 14. Expression of CLASP-2 upon T-cell activation as assayed by Northern analysis. Jurkat cells were activated using PMA, Ionomycin, and αCD28. RNA was prepared from cell culture aliquots at 0 , 1, 2, 4, 8, 14 hours post activation and Northern analysis was performed (A). Hybridization signals obtained with a CLASP-2-specific probe were quantified using a phosphor imager system. Relative signal intensities (refers to total signal of the specific probe used) are shown in the bar diagram (B). The ethidium staining of the Northern gel (A) demonstrates even RNA loading.

DETAILED DESCRIPTION

5.0 Definitions

Except when noted, the terms "patient" or "subject" are used interchangeably and refer to mammals such as human patients and non-human primates, as well as experimental animals such as rabbits, rats, and mice, and other animals.

The term "biological sample" as used herein is a sample of biological tissue, fluid, or cells that contains hCLASP-2 or nucleic acid encoding hCLASP-2 protein. Such samples include, but are not limited to, tissue isolated from humans. Biological samples may also include sections of tissues such as frozen sections taken for histologic purposes. A biological sample is typically obtained from a eukaryotic organism, preferably eukaryotes

such as fungi, plants, insects, protozoa, birds, fish, reptiles, and preferably a mammal such as rat, mice, cow, dog, guinea pig, or rabbit, and most preferably a primate such as chimpanzees or humans.

The term "treating" includes the administration of the compounds or agents of the present invention to prevent or delay the onset of the symptoms, complications, or biochemical indicia of a disease, alleviating the symptoms or arresting or inhibiting further development of the disease, condition, or disorder (*e.g.*, autoimmune disease). Treatment may be prophylactic (to prevent or delay the onset of the disease, or to prevent the manifestation of clinical or subclinical symptoms thereof) or therapeutic suppression or alleviation of symptoms after the manifestation of the disease.

The term "lymphocyte" as used herein has the normal meaning in the art, and refers to any of the mononuclear, nonphagocytic leukocytes, found in the blood, lymph, and lymphoid tissues, *i.e.*, B and T lymphocytes.

The terms "isolated," or "purified," refer to material that is substantially free from components that normally accompany it as found in its native state (*e.g.*, recombinantly produced or purified away from other cell components with which it is naturally associated). Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The terms "nucleic acid" and "polynucleotide" are used interchangeably" and refer to refers to DNA, RNA and nucleic acid polymers containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The amino acids may be natural amino acids, or include an artificial chemical mimetic of a corresponding naturally occurring amino acid.

As used herein a "nucleic acid probe" is defined as a nucleic acid capable of specifically binding to a target nucleic acid of complementary sequence (*e.g.*, through complementary base pairing). As used herein, a probe may include natural (*i.e.*, A, G, C, or T) or modified bases (7-deazaguanosine, inosine, and the like). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization (*e.g.*, probes may be peptide nucleic acids). The probes can be directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind.

The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or, in the case of cells, to progeny of a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

The term "sequence identity" refers to a measure of similarity between amino acid or nucleotide sequences, and can be measured using methods known in the art, such as those described below:

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region (see, *e.g.*, SEQ ID NO: 1), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least of at least 60%, often at least 70%, preferably at least 80%, most preferably at least 90% or at least 95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 bases or residues in length, more preferably over a region of at least about 100 bases or residues, and most preferably the sequences are substantially identical over at least about 150 bases or residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

The phrase "sequence similarity" in the context of two nucleic acids or polypeptides, refers to two or more sequences that are identitical or in the case of amino acids, have homologous amino acid substitutions at either 50%, often at least 60%, often at least 70%, preferably at least 80%, most preferably at least 90% or at least 95% of the indicated positions.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters. For sequence comparison of nucleic acids and proteins to CLASP-2 nucleic acids and proteins, the BLAST and BLAST 2.0 algorithms and the default parameters discussed below are used.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, 1981, Adv. Appl. Math. 2: 482), by the homology alignment algorithm of Needleman & Wunsch, 1970, J. Mol. Biol. 48: 443, by the search for similarity method of Pearson & Lipman, 1988, Proc. Natl. Acad. Sci. U.S.A. 85: 2444, by computerized implementations of these algorithms

(FASTDB (Intelligenetics), BLAST (National Center for Biomedical Information), GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, *e.g.*, Ausubel *et al.*, 1987 (1999 Suppl.), Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y.)

A preferred example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the FASTA algorithm, which is described in Pearson, W.R. & Lipman, D.J., 1988, Proc. Natl. Acad. Sci. U.S.A. 85: 2444. See also W. R. Pearson, 1996, Methods Enzymol. 266: 227-258. Preferred parameters used in a FASTA alignment of DNA sequences to calculate percent identity are optimized, BL50 Matrix 15: -5, k-tuple= 2; joining penalty= 40, optimization= 28; gap penalty -12, gap length penalty =-2; and width= 16.

Another preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, 1977, Nuc. Acids Res. 25: 3389-3402 and Altschul *et al.*, 1990, J. Mol. Biol. 215: 403-410, respectively. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http: //www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences)

uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989, Proc. Natl. Acad. Sci. U.S.A. 89: 10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, *e.g.*, Karlin & Altschul, 1993, Proc. Natl. Acad. Sci. U.S.A. 90: 5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

Another example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, 1987, J. Mol. Evol. 35: 351-360. The method used is similar to the method described by Higgins & Sharp, 1989, CABIOS 5: 151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, *e.g.*, version 7.0 (Devereaux *et al.*, 1984, Nuc. Acids Res. 12: 387-395.

Another preferred example of an algorithm that is suitable for multiple DNA and amino acid sequence alignments is the CLUSTALW program (Thompson, J. D. *et al.*,

1994, Nucl. Acids. Res. 22: 4673-4680). ClustalW performs multiple pairwise comparisons between groups of sequences and assembles them into a multiple alignment based on homology. Gap open and Gap extension penalties were 10 and 0.05 respectively. For amino acid alignments, the BLOSUM algorithm can be used as a protein weight matrix (Henikoff and Henikoff, 1992, Proc. Natl. Acad. Sci. U.S.A. 89: 10915-10919).

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available (*e.g.*, the polypeptide of SEQ ID NO: 1 can be made detectable, *e.g.*, by incorporating a radiolabel into the peptide, and used to detect antibodies specifically reactive with the peptide).

The term "sorting" in the context of cells as used herein to refers to both physical sorting of the cells, as can be accomplished using, *e.g.*, a fluorescence activated cell sorter, as well as to analysis of cells based on expression of cell surface markers, *e.g.*, FACS analysis in the absence of sorting.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

The phrase "specifically (or selectively) binds" to an antibody refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample.

The phrase "specifically bind(s)" or "bind(s) specifically" when referring to a peptide refers to a peptide molecule which has intermediate or high binding affinity, exclusively or predominately, to a target molecule. The phrases "specifically binds to" refers to a binding reaction which is determinative of the presence of a target protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated assay conditions, the specified binding moieties bind preferentially to a particular target protein and do not bind in a significant amount to other components present in a test sample. Specific binding to a target protein under such conditions may require a binding moiety that is selected for its specificity for a particular target antigen. A variety of assay formats may be

used to select ligands that are specifically reactive with a particular protein. For example, solid-phase ELISA immunoassays, immunoprecipitation, Biacore and Western blot are used to identify peptides that specifically react with PDZ domain-containing proteins. Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 times background. Specific binding between a monovalent peptide and a PDZ-containing protein means a binding affinity of at least 10^4 M^{-1}, and preferably 10^5 or 10^6 M^{-1}.

The phrase "homotypic interaction" refers to the binding of a given protein to another molecule of the same protein (*e.g.*, the binding of hCLASP-2 to hCLASP-2). The phrase "heterotypic interaction" refers to the binding of a given protein to a different protein or other molecule (*e.g.*, the binding of hCLASP-2 to a PDZ domain-containing protein or the binding of a transcription factor to DNA).

The phrase "immune cell response" refers to the response of immune system cells to external or internal stimuli (*e.g.*, antigen, cytokines, chemokines, and other cells) producing biochemical changes in the immune cells that result in immune cell migration, killing of target cells, phagocytosis, production of antibodies, other soluble effectors of the immune response, and the like.

The terms "B lymphocyte response" and "B lymphocyte activity" are used interchangeably to refer to the component of immune response carried out by B lymphocytes (*i.e.* the proliferation and maturation of B lymphocytes, the binding of antigen to cell surface immunogobulin, the internalization of antigen and presentation of that antigen via MHC molecules to T lymphocytes, and the synthesis and secretion of antibodies).

The terms "T lymphocyte response" and "T lymphocyte activity" are used here interchangeably to refer to the component of immune response dependent on T lymphocytes (*i.e.*, the proliferation and/or differentiation of T lymphocytes into helper, cytotoxic killer, or suppressor T lymphocytes, the provision of signals by helper T lymphocytes to B lymphocytes that cause or prevent antibody production, the killing of specific target cells by cytotoxic T lymphocytes, and the release of soluble factors such as cytokines that modulate the function of other immune cells).

The term "immune response" refers to the concerted action of lymphocytes, antigen presenting cells, phagocytic cells, granulocytes, and soluble macromolecules produced by the above cells or the liver (including antibodies, cytokines, and complement) that results in selective damage to, destruction of, or elimination from the human body of

invading pathogens, cells or tissues infected with pathogens, cancerous cells, or, in cases of autoimmunity or pathological inflammation, normal human cells or tissues.

Components of an immune response may be detected in vitro by various methods that are well known to those of ordinary skill in the art. For example, (1) cytotoxic T lymphocytes can be incubated with radioactively labeled target cells and the lysis of these target cells detected by the release of radioactivity, (2) helper T lymphocytes can be incubated with antigens and antigen presenting cells and the synthesis and secretion of cytokines measured by standard methods (Windhagen A; *et al.*, 1995, Immunity 2(4): 373-80), (3) antigen presenting cells can be incubated with whole protein antigen and the presentation of that antigen on MHC detected by either T lymphocyte activation assays or biophysical methods (Harding *et al.*, 1989, Proc. Natl. Acad. Sci., 86: 4230-4), (4) mast cells can be incubated with reagents that cross-link their Fc-epsilon receptors and histamine release measured by enzyme immunoassay (Siraganian, *et al.*, 1983, TIPS 4: 432-437).

Similarly, products of an immune response in either a model organism (*e.g.*, mouse) or a human patient can also be detected by various methods that are well known to those of ordinary skill in the art. For example, (1) the production of antibodies in response to vaccination can be readily detected by standard methods currently used in clinical laboratories, *e.g.*, an ELISA; (2) the migration of immune cells to sites of inflammation can be detected by scratching the surface of skin and placing a sterile container to capture the migrating cells over scratch site (Peters *et al.*, 1988, Blood 72: 1310-5); (3) the proliferation of peripheral blood mononuclear cells in response to mitogens or mixed lymphocyte reaction can be measured using ^{3}H-thymidine; (4) the phagocitic capacity of granulocytes, macrophages, and other phagocytes in PBMCs can be measured by placing PMBCs in wells together with labeled particles (Peters *et al.*, 1988); and (5) the differentation of immune system cells can be measured by labeling PBMCs with antibodies to CD molecules such as CD4 and CD8 and measuring the fraction of the PBMCs expressing these markers.

As used herein, the phrase "signal transduction pathway" or "signal transduction event" refers to at least one biochemical reaction, but more commonly a series of biochemical reactions, which result from interaction of a cell with a stimulatory compound or agent. Thus, the interaction of a stimulatory compound with a cell generates a "signal" that is transmitted through the signal transduction pathway, ultimately resulting in a cellular response, *e.g.*, an immune response described above.

A signal transduction pathway refers to the biochemical relationship between a variety of signal transduction molecules that play a role in the transmission of a signal from

one portion of a cell to another portion of a cell. Signal transduction molecules of the present invention include, for example, extracellular and intracellular domains of CLASP-2. As used herein, the phrase "cell surface receptor" includes molecules and complexes of molecules capable of receiving a signal and the transmission of such a signal across the plasma membrane of a cell. An example of a "cell surface receptor" of the present invention is the T cell receptor (TCR). As used herein, the phrase "intracellular signal transduction molecule" includes those molecules or complexes of molecules involved in transmitting a signal from the plasma membrane of a cell through the cytoplasm of the cell, and in some instances, into the cell's nucleus. In the present invention, CLASP-2 can be referred to as an "intracellular signal transduction molecule", but can also be referred to as a "signal transduction molecule".

A signal transduction pathway in a cell can be initiated by interaction of a cell with a stimulator that is inside or outside of the cell. If an exterior (*i.e.*, outside of the cell) stimulator (*e.g.*, an MHC-antigen complex on an antigen presenting cell) interacts with a cell surface receptor (*e.g.*, a T cell receptor), a signal transduction pathway can transmit a signal across the cell's membrane, through the cytoplasm of the cell, and in some instances into the nucleus. If an interior (*e.g.*, inside the cell) stimulator interacts with an intracellular signal transduction molecule, a signal transduction pathway can result in transmission of a signal through the cell's cytoplasm, and in some instances into the cell's nucleus.

Signal transduction can occur through, *e.g.*, the phosphorylation of a molecule; non-covalent allosteric interactions; complexing of molecules; the conformational change of a molecule; calcium release; inositol phosphate production; proteolytic cleavage; cyclic nucleotide production and diacylglyceride production. Typically, signal transduction occurs through phosphorylating a signal transduction molecule. According to the present invention, a CLASP-2 signal transduction pathway refers generally to a pathway in which CLASP-2 protein regulates a pathway that includes engaged-receptors, PKC-substrates, G proteins, and other molecules.

5.1. Introduction

The present invention relates to a novel transmembrane protein, CLASP-2, a new member of the CLASP family that contains an endodomain that displays the appropriate properties to organize the cytoskeleton and signal transduction apparatus of the immune gateway.

CLASP-2 functions in cells of the immune system, *e.g.*, T cells and B cells, as well as non-immune cells. The CLASP-2 protein functions in a variety of cellular processes,

particularly related to immune function, regulation of T cell and B cell interactions, T cell activation, and in the organization, establishment and maintenance of the "immunological synapse" (see Dustin *et al.*, 1999, Science 283: 680-682; Paul *et al.*, 1994, Cell 76: 241-251; Dustin *et al.*, 1996, J. Immunol. 157: 2014; Dustin *et al.*, 1998, Cell 94: 667), including signal transduction, cytoskeletal interactions, and membrane organization.

Without intending to be bound by a particular mechanism or limited in any way, the CLASP-2 protein is believed to be a component of the lymphocyte organelle called the "immune gateway" that creates a docking site or portal for cell-cell contact during antigen-presentation. It is believed the cytoplasmic domains of CLASP-2 proteins organize it into a patch at the leading edge of T cells. The carboxy-terminus encoded sequences mediate interaction with PDZ domain proteins and with cytoskeletal proteins (*e.g.*, spectrin or ankyrin) to connect CLASP-2 to the microtubule network and hold the receptors at a polarized configuration just above the microtubule-organizing center ("MTOC"). Thus, when T cells engages a B cell acting as an APC, the CLASP-2 molecules engage one another to dock the two cells and organize the immune synapse.

Modulating the expression of the CLASP-2 protein, and interference with, or enhancement of, CLASP-2 protein interactions with other proteins has a number of beneficial physiological effects, *e.g.*, altered signaling in response to antigen, altered T and B cell response to antigen, and modulation of T cell activation. In one aspect, the CLASP-2 extracellular domain is targeted (*e.g.*, using anti-CLASP-2 antibody, soluble CLASP-2 fragments, and the like) to regulate T cell activation (and thus regulate immune responses). Disorders that can be treated by disrupting CLASP-2 function, include without limitation, multiple sclerosis, juvenile diabetes, rheumatoid arthritis, pemphigus, pemphigoid, epidermolysis bullosa acquista, lupus, endometriosis, toxemia or pregnancy induced hypertension, pruritic urticarial papules and plaques of pregnancy (PUPPP), herpes gestationis, impetigo herpetiformis, pruritus gravidarum, placenta-related disorders, and Rh incompatibility.

In another aspect, the present invention provides methods and reagents for detection of CLASP-2 expression and CLASP-2-expressing cells. Abnormal expression patterns or expression levels are diagnostic for immune and other disorders. For example, diseases characterized by overproduction or depletion of lymphocytes in blood or other organs may be detected or monitored by monitoring the level of CLASP-2 polypeptide or mRNA in a biological sample (*e.g.*, peripheral blood), *e.g.*, the number or percentage of CLASP-2 expressing cells. Diseases characterized by overproduction of T cells include, *e.g.*,

leukemia (both ALL and CLL), lymphoma (including non-Hodgkins lymphoma, Burkitt's lymphoma, mycosis fungoides, and sezary syndrome), EBV, CMV, toxoplasmosis, syphilis, typhoid, brucellosis, tuberculosis, influenza, hepatitis, serum sickness, and thyrotoxicosis. Diseases associated with the depletion of T cells include, *e.g.*, HIV and myelodysplasia. Diseases associated with the overproduction of B cells include, *e.g.*, leukemia (both ALL and CLL), non-Hodgkins lymphoma, Burkitt's lymphoma, myeloma, EBV, CMV, toxoplasmosis, syphilis, typhoid, brucellosis, tuberculosis, influenze, hepatitis, serum sickness, and thyrotoxicosis. Diseases associated with the depletion of B cells include, *e.g.*, myelodysplasia.

5.2. CLASP-2 cDNA and Polypeptide Structure

The CLASP-2 protein is type I transmembrane glycoprotein, characterized by multiple forms produced by alternative exon usage (*i.e.*, production of splice variants). In one naturally occurring form, CLASP-2 has the structure shown in FIG. 1. However, as discussed in detail *infra*, the CLASP-2 gene encodes a variety of gene product due to alternative splicing of mRNA. FIG. 2 shows the nucleotide sequence and conceptual translation of human CLASP-2 polypeptides:

hCLASP-2A cDNA (SEQ ID NO: 1) and hCLASP-2A polypeptide (SEQ ID NO: 2).

hCLASP-2B cDNA (SEQ ID NO: 3) and hCLASP-2B polypeptide (SEQ ID NO: 4).

hCLASP-2C cDNA (SEQ ID NO: 5) and hCLASP-2C polypeptide (SEQ ID NO: 6).

hCLASP-2D cDNA (SEQ ID NO: 7) and hCLASP-2D polypeptide (SEQ ID NO: 8).

hCLASP-2E cDNA (SEQ ID NO: 9) and hCLASP-2E polypeptide (SEQ ID NO: 10).

Unless specifically referred to, the phrase "human CLASP-2 (hCLASP-2)" is used herein refers to hCLASP-2A, hCLASP-2B, hCLASP-2C and hCLASP-2E. "hCLASP-2D" cDNA is also known as KIAA1058, which was described by Kikuno *et al.*, 1999, *DNA Res.* 6, 197-205 as a cDNA from brain encoding a protein of unknown function.

CLASP-2 polypeptides typically include an approximately 120 residue leader sequence, followed by a cadherin proteolytic cleavage signal RXXR, an extracellular domain, a transmembrane domain, and an intracellular domain. The present invention provides a polynucleotide having the sequence of SEQ. ID. NO: 1, or a fragment thereof, and a polypeptide having the sequence of SEQ. ID NO: 2, or a fragment thereof. In addition, the

invention provides polynucleotides comprising hCLASP-2 genomic sequences, CLASP-2 homologs from other species, naturally occurring alleles of hCLASP-2, and hCLASP-2 variants as described herein, and methods for using CLASP-2 polynucleotide, polypeptides, antibodies and other reagents.

5.2.1. CLASP-2 Polypeptide Domains

As is shown in FIG. 1, one naturally occurring CLASP-2 cDNA encodes a polypeptide characterized by several structural and functional domains and defined sequence motifs. To provide guidance to the practitioner, the structural features are described *infra*. However, it will be understood that the present invention is not limited to polypeptides that include all, or any particular one of these domains or motifs. For example, a CLASP-2 fusion protein of the invention contains only the extracellular domain of CLASP-2. Similarly, the CLASP-2A polypeptide of SEQ ID NO: 2 does not have the ITAM motifs (discussed *infra*) found in the CLASP-2B and 2C polypeptides.

It will be appreciated that the structurally (and functionally) different domains of CLASP-2 polypeptides (and the corresponding region of the mRNA) are of interest, in part, because they may be separately targeted or modified (*e.g.*, deleted or mutated) to affect the activity or expression of a CLASP-2 gene product (in order to, for example, modulate an immune response). For example, the extracellular domain of a CLASP-2 protein can be targeted (*e.g.*, using an anti-CLASP monoclonal antibody to (a) block the interaction of a CLASP-2-expressing cell (*e.g.*, a T cell) and a second cell (*e.g.*, a B cell) displaying a protein that is bound by CLASP-2 (*i.e.*, a CLASP-2 ligand). Similarly, an intracellular domain (*e.g.*, ITAM or DOCK, see *infra*) can be targeted to interfere with signal transduction without interfering with extracellular ligand binding.

Generally, inhibiting CLASP-2 expression or CLASP-2 polypeptide function will result in modulation of immune function including, for example, changing the threshold for T cell activation by affecting formation of the immune synapse. Modulation of immune function can be screened and quantitated by a number of assays known in the art and described herein (see also §5.14).

5.2.1.1. Signal Peptide

The human CLASP-2 sequence presented in FIG. 1 encodes two potential start sites for translation. The first predicted methionine appears at nucleotide 278 (ATG). The second methionine appears at nucleotide 476. Both have an acceptable consensus sequence

for a translational start (A/GxxATGG; Kozak, M., 1996, Mamm. Genome 7(8): 563-74). A polypeptide beginning at the second methionine is also predicted to encode a signal peptide capable of localizing the protein to the secretory pathway by SignalP, a signal sequence prediction program (Nielsen, H. *et al.*, 1997, Protein Eng. 10(1): 1-6). Polypeptides beginning at the first methionine are not predicted to contain a signal sequence; however, the consensus for signal prediction is only 80-90% accurate for known signal sequences. A third possibility for a translational start is that the cDNA listed in FIG. 1 is incomplete and another methionine is encoded in frame and upstream of the sequence shown in FIG.1 .

5.2.1.2. Extracellular Domain

The CLASP-2 extracellular domain is characterized by one cadherin EC-like motif (Pigott, R. and Power, C., 1993, The Adhesion Molecule Factbook. Academic Press, pg. 6; Jackson, R. M. and Russell, R. B., 2000, J. Mol. Biol. 296: 325-34). Several highly conserved cysteines are found in the extracellular domain, as well as various glycosylation signals. Through its extracellular domains, CLASP-2 may interact with ligands in a homotypic and/or heterotypic manner to establish the immunological synapse in conjunction with molecules such as TCR, MHC class I, MHC class II, CD3 complex and accessory molecules such as CD4, CD3, ICAM-1, LFA-1, and others. Many cadherins contain a pro-domain of approximately 50 to 150 amino acids that is removed before localization to the plasma membrane. This cleavage is presumed to be carried out by Furin (Posthaus, H. *et al.*, 1998, FEBS Let 438: 306-10) at a consensus sequence of RKQR. Furin is a protease that is at least partially responsible for the maturation of certain cadherins. CLASP-2 has the sequence RNQR at nucleotides 945 through 957. By homology, this region is around 120 amino acids into the predicted protein start site for hCLASP-2A. This region may be a pro-domain and cleavage may be required for CLASP-2 function, or aspects of CLASP-2 function.

Antibodies raised against the extracellular domain can be added to cells expressing CLASP-2. These antibodies can either block the interaction of CLASP-2 with potential ligands or stabilize these interactions. Any immunoassay known in the art, *e.g.*, listed and described herein, may be used to assess the modulation of immune function brought about by this approach.

Similarly, portions of the extracellular domain of CLASP-2 can be expressed as soluble protein. This soluble protein can then be added to cells expressing CLASP-2. These proteins may interact with potential ligands to competitively inhibit their binding to endogenous CLASP-2. This could modulate CLASP-2 function via the immunoassays

described herein. Recombinant proteins could interfere in a positive or negative fashion with CLASP-2 interactions.

5.2.1.3. Transmembrane Domain

CLASP-2 predicted amino acid sequence was analyzed using the PHDhtm analysis software for prediction of transmembrane helices (Rost, B., *et al.*, 1996, Prot. Science 7: 1704-1718). Using the PPHDhtm analysis software, it was determined that the transmembrane domain is located from nucleotides 2861-2917 (see FIG. 1), as well as three other potential transmembrane domains located near the amino terminal end.

5.2.1.4. Intracellular Domains

The CLASP-2 intracellular domains contain motifs corresponding to several types of protein domains. Depending on the specific CLASP-2 (*i.e.*, specific family member or splice variant) all or only some of the domains can be present. Listed from amino terminus to carboxy terminus, the domains include: (1) ITAM (Chan *et al.* 1994, *Annual Review of Immunology* 12: 555-592), (2) a newly discovered DOCK/CLASP-2 motif, (3) a coiled-coil motif, and (4) a C-terminal PDZ binding motif (PBM) (also referred to as PDZ ligand or "PL").

5.2.1.5. ITAM

Immunoreceptor Tyrosine-based Activation Motifs (ITAM motifs; also known as ARAM, or antigen recognition activation motifs) are motifs contained within antigen receptors for T and B cells, and Fc receptors on other leukocytes, and are necessary for proper activation and signal transduction in these cells. They are characterized by the consensus sequence YXXL/I - X7/8- YXXL/I (Grucza *et al.*, 1999, Biochemistry 38: 5024-5033), usually separated by 6-8 amino acids (Watson *et al.*, 1998, Immunol. Today 19: 260-264; Isakov, J. Leukoc. Biol. 61: 6-16). ITAM is used as an intracellular regulatory motif through its ability to be tyrosine phosphorylated by src-family tyrosine kinases such as Lyn that are involved in leukocyte signal transduction. Once phosphorylated, the ITAM acts as a high affinity binding site for SH2 containing proteins. Signal transduction components including ZAP-70, Syk, Lyn, Shc, PI3 kinase, and Grb2 contain SH2 domains and have been shown to bind ITAMs (Clements *et al.*, 1999, Annu. Rev. Immunol. 17: 89-108). This places ITAM-containing molecules in a central role of intracellular signal regulation in leukocytes. ITAM motifs in leukocyte signaling can facilitate signal transduction (*e.g.*,

tyrosine kinase signaling) by acting as temporal scaffolds where other transduction components could bind and be properly positioned to mediate transduction. ITAM motifs often appear in multiples in a protein, however, it is known that one set of YXXL/I alone can transduce signals of the PTK pathway, though weakly.

CLASP-2 proteins typically have ITAM YXXL/I motifs (where X is any amino acid) separated by 3 or 13 amino acids. In various embodiments the CLASP-2 polypeptide of the invention is characterized by one or more of the motifs shown in Table 1.

Table 1

CLASP-2 ITAM Motifs

Motif No.	Sequence Motif
1	YXX(I/L)-X_3-YXX(I/L)
2	YXX(I/L)-X_{13}-YXX(I/L)
3	YXX(I/L)-X_3-YXX(I/L)-X_{13}-YXX(I/L)

The presence of multiple ITAM motifs in CLASPs proteins indicates that they may be engaged by multiple signal transduction components (*e.g.*, ZAP-70/Syk, Shc, PI3 kinase, and Grb2). In general, the ITAM motif in CLASP proteins match identically to the canonical ITAM motif with some motifs containing a conservative amino acid change (*i.e.* valine instead of isoleucine or leucine). As previously described for other ITAMs, the ITAMs within CLASPs can bind SH2-containing proteins including ZAP-70, Syk, Shc, PI3 kinase, and Grb2. Since CLASPs have an extracellular domain, CLASPs protein can independently initiate a signal transduction cascade through engagement of its extracellular domain. Otherwise CLASPs may cooperate with an antigen receptor signaling complex (*e.g.*, with CD3/TCR, BCR, FcR), to facilitate tyrosine kinase signal transduction

The ITAMs have demonstrated different binding specificity and affinities for SH2 domains (Clements, *et al.*, 1999, Ann. Rev. Immunol. 17: 89-108). For example, Shc, PI3 kinase, and Grb2 bind to dual and mono phosphorylated ITAMs with different affinities. Thus the ITAMs in CLASPs are believed to provide quantitative as well as qualitative differences in signal transduction depending up their phosphorylation state, as well as to inhibit or augment specific protein interactions and hence specific tyrosine kinase-mediated signaling pathways in leukocytes.

Antagonizing the PTK-CLASP-2 interaction (*e.g.*, phosphorylation of CLASP-2) will thus inhibit immune function. In one embodiment, interactions between ITAM-bearing human CLASPs and their binding partners are believed to be antagonized by the alpha subtype (SIRPalpha) of signal regulatory proteins that has been shown to negatively

regulate ITAM-dependent lymphocyte activation (Lienard H; 1999, J Biol Chem 274: 32493-9). Also, a recently recognized family of immunoreceptor tyrosine-based inhibition motif (ITIM) receptors are thought to inhibit the ITAM-induced activation of immune competent cells (Gergely, *et al.*, 1999, J. Immunol Lett 68: 3-15) and therefore may block CLASP-partner interaction.

5.2.1.6. DOCK

CLASP-2 polypeptides contain a new "DOCK" motif, not previously described in the scientific literature. The CLASP DOCK motif includes a series of five tyrosines surrounded by conserved sequences in regions A, B, C, D, and G (see FIG. 5B). There are also two highly conserved non-tyrosine containing regions (E and G) separated by nine amino acids (P+EXAI+XM) and (LXMXL+GXVXXXVNXG) (where X is any amino acid).

The cytoplasmic region of CLASP-2 immediately following the ITAM domains exhibits sequence similarity to the C-terminal third of the so-called "DOCK" proteins. The DOCK gene family includes three molecules that are the human homologues of the *C. elegans* CED proteins known to be involved in apoptosis. CED-5 (DOCK180), a major CRK-binding protein, alters cell morphology upon translocation to the membrane (mediates the membrane motion that scavenger cells exhibit as they surround and engulf dying cells; its function can be partially rescued by the human DOCK180 (Wu *et al.*, 1998, *Nature* 392: 501-504). Myoblast City in Drosophila (MBC) is another member of the DOCK protein family and has been found to be involved in myoblast fusion (Erickson, *et al.*, 1997, J. Cell Biol. 138: 589). Since CLASP-2 expression is found in syncytial tissues such as placenta, muscle, and heart, it is believed that CLASP-2 is involved in mediating or inhibiting cell fusion.

The DOCK family has been implicated in the control of cell shape. DOCK1, when transfected into spindle cells, can make them flattened and polygonal (Takai, *et al.*, 1996, Genomics 35: 403-303). DOCK1 expression is ubiquitous except in hematopoetic cells. DOCK2 is expressed in hematopoetic cells and when transfected into spindle cells can make them round up (Nishihara, H., 1999, Hokkaido Igaku Zasshi 74: 157-66). DOCK2 is expressed in peripheral blood lymphocytes, thymus, spleen, and liver.

5.2.1.7. COILED-COIL

CLASP-2s have the two coiled-coil domains (Lupas *et al.*, 1991, *Science* 252: 1162-64; Lupas, A., 1996, Meth. Enzymology 266: 513-525). Coiled-coil domains are known to interact directly with cytoskeleton, indicating that that CLASP-2 proteins interact directly with the cytoskeleton. Thus, it is believed that CLASP-2 binds cytoskeletal proteins, *e.g.*, spectrin, ankyrin, hsp70, talin, ezrin, tropomyosin, myosin, plectin, syndecans, paralemmin, Band 3 protein, Cytoskeletal protein 4.1, Tyrosine phosphatase PTP36 and other molecules.

5.2.1.8. PDZ Ligand

CLASP-2 proteins contain a PDZ-ligand motif ("PBM" or "PL") at the C-terminus of the protein. This short (3 – 8 amino acid) motif mediates the binding of proteins terminating at their carboxyl terminus in the motif (most commonly S/T – X – V – free carboxyl-terminus) to other proteins containing one or more specific PDZ domains (See Songyang *et al.*, 1997, Science 275: 72 and Doyle *et al.*, 1996, Cell 85: 1067 for a discussion of PDZ-ligand structures).

PDZ domain-containing proteins are involved in the organization of ion channels and receptors at the neurological synapse and in establishing and maintaining polarity in epithelial cells via their binding to the C-termini of transmembrane receptors. It has been shown that PDZ-domain containing proteins can mediate protein-protein interactions in immune system cells (*e.g.*, DLG1 binds to the lymphocyte potassium channel KV1.3 in human T lymphocytes, (Hanada *et al.*, 1997, J. Biol. Chem. 272: 26899).

Biochemical evidence that CLASP-2 interacts with the PDZ domains of three closely related proteins is shown in FIG 9A-D. FIG. 9A demonstrates the specificity of the interaction, as the C-terminal 20 amino acids of CLASP-2 bind PSD-95, NeDLG, and DLG1, but not to the PDZ domains of the TIAM-1 protein. FIG. 9B demonstrates the affinity of the interaction. Notably, the highest affinity interaction occurs between CLASP-2 and NeDLG, with a specific binding affinity of at least 10^4 M^{-1}. Affinities in the micromolar range have been found for other biologically important PDZ-ligand interactions. FIG. 9C demonstrates the ability to inhibit CLASP-2 PDZ interactions using either a short fragment of CLASP-2 (the eight C-terminal amino acids) or the C-terminus of KV1.3. As noted above, KV1.3 is known to bind to DLG1 in live lymphocytes. FIG. 9D demonstrates that CLASP-2 and KV1.3 compete for PDZ binding; *i.e.*, not only does KV1.3 block CLASP-2 binding but

CLASP-2 also blocks KV1.3 binding. The ability of the eight C-terminal residues of CLASP-2 to inhibit the interaction of both CLASP-2 and KV1.3 with selected PDZ domains suggests that compounds related to the C-terminal eight-amino acids of CLASP-2, when introduced into cells, will mediate changes in multiple protein-protein interactions involved in the function of lymphoid tissues and other tissues that express these proteins (including heart, lung, and kidney).

Evidence that the C-terminal 8 amino acids of CLASP-2, when introduced into cells, can effect cellular function comes from the experiments in which these amino acids were introduced into cells as a fusion, *e.g.*, with the HIV-derived TAT transporter peptide sequence. Addition of the TAT-CLASP-2 fusion peptide to Jurkat T lymphocytes (compared to controls using the TAT peptide alone) results in subtle, time-dependent alterations in intracellular calcium concentrations as measured using the calcium indicator dye Fluo-4. While these results are consistent with the hypothesis that the TAT-CLASP-2 fusion changes T cell ion fluxes. In particular, the results indicate that the CLASP-2 C-terminal sequence can slightly increase basal intracellular calcium concentrations and can slightly decrease the proportional increase in calcium upon activation of the cells with anti-CD3 antibody. Such changes would be expected for a compound that disrupts localization of the T cell activation-associated CLASP-2 protein and the KV1.3 potassium channel. Small changes in T cell calcium flux can result in large changes in the functional activity of the cells (Wulfing *et al.*, 1997, J. Exp. Med. 185: 1815).

5.2.1.9. Modulation of Immune Responses

CLASP-2 proteins, as described above, modulate immune function in a variety of ways and through a variety of mechanisms (*i.e.*, changing the threshold for T cell activation) by affecting formation of the immunological synapse. Establishment and maintenance of the immunological synapse can involve: (A) signal transduction, (B) cell-cell interactions, and (C) membrane organization.

(A) Signal transduction

Human CLASP proteins, as discussed above, contain SH3 domains and tyrosine phosphorylation sites. These regions have been shown to be involved in signal transduction in a variety of cells including lymphocytes. Thus, human CLASP proteins are believed to interact with these regions during signal transduction events which lead to modulation of immune responses.

CLASP proteins can interact with Tec sub-family of nonreceptor tyrosine kinases. The Tec sub-family of nonreceptor tyrosine kinases consists of Tec, Btk, Tsk/Itk/Emt Itk, and Bmx, and is defined by the presence of SH3 and SH2 domains adjacent to the catalytic domain and an amino-terminal region containing a pleckstrin homology (PH) domain, a Tec homology (TH) domain, and a proline-rich region (Mano, H.; 1999, Cytokine Growth Factor Rev 10: 267-80). The T cell specific Tsk/Itk/Emt, and Btk expressed in most hematopoietic cells other than T cells are important components of antigen receptor signaling pathways in hematopoietic cells.

Btk has been identified as the gene defective in murine X-linked immunodeficiency (*xid*) and human X-linked agammaglobulinemia (XLA) (Nisitani, S., 2000, Proc Natl Acad Sci U.S.A. 97: 2737-42). In *xid* mice, B cell numbers are reduced to one-half of normal and the titers of specific immunoglobulin isotypes are significantly reduced; in addition, *xid* B cells are insensitive to a number of mitogenic stimuli. The human disorder is much more severe, resulting in nearly complete elimination of the B cell compartment and dramatically reduced immunoglobulin levels. Biochemical studies have supported multiple roles for Btk in B cell activation. Btk kinase activity and tyrosine phosphorylation are increased after cross-linking either the B cell receptor on B cells or the high affinity IgE receptor, FcRI, on mast cells. Interleukin-5 and interleukin-6 treatment have also been shown to lead to the activation of Btk.

Itk, like Btk, is tyrosine-phosphorylated upon antigen receptor cross-linking (Mano, H., 1999, Cytokine Growth Factor Rev, 10: 267-80). In addition, peripheral T cells from mice lacking functional Itk are refractory to stimulation by antibodies to CD3 plus antigen presenting cells. These Itk-deficient T cells can be stimulated by phorbol ester and calcium ionophore, demonstrating that Itk acts in signaling pathways proximal to the TCR.

Unlike the related Src family tyrosine kinases including Lyn, Lck, Fyn, ZAP-70, SyK, and CSK, the Tec family kinases lack the amino-terminal myristylation site crucial for the membrane localization of Src family kinases, suggesting that some adaptor proteins are required for the their membrane localization (Mano, H., 1999, Cytokine Growth Factor Rev 10: 267-80). Since all the Tec family kinases contain a proline-rich region which could be bound by a SH3 domain, and since all the human CLASPs contain a SH3 domain, it is believed that human CLASPs could serve as adaptors for the members in the Tec family in different hematopoietic cells.

GTP-binding proteins play an important role in immune response (Mach, B., 1999, Science 285: 1367). A number of biochemical events triggered by TCR/CD3-induced

T cell activation are ablated by agents that modulate the action of G proteins. Pertinent to this is the ability of cholera toxin to inhibit the cellular proliferation and intracellular Ca^{2+} mobilization that is mediated by anti-CD3 antibody treatment of T cells. The G protein competitive inhibitor GDPS, can impede the extent of inositol phosphates generated upon stimulation in peripheral T lymphocytes. Nonhydrolyzable analogs of GTP, such as GTPS, or other agents such as ALF that activate G proteins by circumventing the need for receptor engagement, can result in T cell activation.

The Gαq/11subfamily (Stanners, J., 1995, J Biol Chem 270: 30635-42) and Rap1 (Lafont, V., 1998, Biochem Pharmacol 55: 319-24) of GTP-binding proteins have been shown to be involved in human T cell receptor/CD3-mediated signal transduction pathway. Also, Cdc42, a Rho family small GTPase, is known to play a critical role in the formation of actin microspikes in response to external stimuli (Miki, H.; 1998, Nature, 391: 93-6). Interestingly, a Cdc42 binding protein, WASP, has a proline-rich domain which could interact with the SH3 domain present in all the human CLASPs. Human CLASPs may interact with these GTP-binding proteins.

Several adaptor proteins including NCK, CBL (Bachmaier, K., 2000 Nature 403: 211-6), SHC, LNK, SLP-76, HS1, SIT, VAV, GrB2, and BRDG1, and two tyrosine phosphotases, EZRIN, SHP-1 and SHP-2 have been shown to interact with ITAM or SH3 domains. These proteins may also interact with CLASP-2. Several proteins have been shown to interact with ITAM or SH3 domains and may also interact with CLASP-2. These include adaptor proteins such as NCK, CBL (Bachmaier, K., 2000, Nature 403: 211-6), SHC, LAT, LNK, SLP-76 (Krause M *et al.*, 2000, J Cell Biol 149: 181-94), HS1, SIT, VAV, GrB2 (Zhang W. and Samelson, L.E., 2000, Semin Immunol 12: 35-41), and BRDG1, kinases such as SYK and LCK, and tyrosine phosphatases such as SHP-1 and SHP-2. These interactions can be defined by a number of different biochemical or cell biological methods including in vitro binding assays, co-immunoprecipitation assays, co-immunostaining (Harlow, E. and Lane, D., 1999, Using Antibodies: A laboratory Manual. Cold Spring Harbor Press) or genetic assays such as yeast the yeast two hybrid system, in which a CLASP-2 protein or fragment can be used as "bait" (Zervos *et al.*, 1993, Cell 72: 223-232; Madura *et al.*, 1993, J. Biol. Chem 268: 12046-12054).

Other assays include in vitro binding assays, co-immunoprecipitation assays, co-immunostaining assays, and yeast two hybrid system screening assays in which a CLASP-2 domain or fragment can be used as "bait" or "trap" protein (Zervos *et al.* (1993), Cell 72: 223-232; Madura *et al.* (1993) J. Biol. Chem. 268: 12046-12054).

In other embodiments, CLASP polypeptides are transfected into lymphocytes. After transfection, a variety of standard assays can be used to evaluate, for example, CLASP modulation of T cell activation. These assays include calcium influx assays, NF-AT nuclear translocation assays (*e.g.*, Cell, 1998, 93: 851-61), NF-AT/luciferase reporter assays (*e.g.*, MCB 1996 16: 7151-7160), tyrosine phosphorylation of early response proteins such as HS1, PLC-γ, ZAP-76, and Vav (*e.g.*, J. Biol. Chem. 1997, 272: 14562-14570).

(B) Cell-Cell Interaction

As discussed above, human CLASP proteins are homologues of E-cadherin. As shown in FIG. 1, CLASP-2 contains both a cadherin cleavage domain and a cadherin ectodomain. Therefore CLASP-2 proteins may interact with cadherins through these domains. The cadherins constitute a family of cell surface adhesion molecules that are involved in calcium-dependent cell to cell adhesion. Human cadherins, E-, P- N- and VE-cadherin, have a restricted tissue distribution: E- and P-cadherin are expressed in epithelial tissues, N-cadherin is found mainly on neural cells, and VE-cadherin is found on vascular endothelium. Homophilic binding between cadherins on adjacent cells is vital for the maintenance of strong cell to cell adhesion in these tissues. For example E-cadherin is required for the formation of adherens junctions between mature epithelial cells and is involved in Langerhans cell adhesion to keratinocytes, and VE-cadherin is needed for the maintenance of lateral association between endothelial cells. The extracellular regions of mature mammalian cadherins are comprised of five "CAD" modules of approximately 1110 amino acids. Crystallographic and biochemical studies indicate that cadherins can form dimers on the cell surface, and that interaction with dimeric cadherins on opposing cell surfaces can lead to the formation of "zipper-like" cell junctions.

The integrins are a second family of transmembrane adhesion molecules that are involved in both cell to cell and cell to matrix interactions. At least 15 chains associate with 8 chains to form a large number of heterodimeric integrins that can be classified into several major subfamilies based on their shared use of a particular chain. Members of three subfamilies, the 1, 2, and 7 integrins, are commonly found on leukocytes. The expression of 1 integrins is widespread (for example, 51, CD49e/CD29, is found on T cells, granulocytes, platelets, fibroblasts, endothelium, and epithelium), whereas the 2 and 7 integrins have a restricted pattern of expression.

Interestingly, E-cadherin on human epithelial cells has been found to be a ligand for the mucosal lymphocyte integrin, E7, and a similar interaction has been indicated

in the mouse. Monoclonal antibodies to E-cadherin or to E7 block IEL adherence to epithelial cells, and transfection of cells with E7 confers upon them the ability to adhere to cells transfected with E-cadherin.

L929 cells can be transfected with CLASP-2 and Neomycin. G418-resistant clones can be screened for CLASP-expression with anti-CLASP peptide-specific antibodies. CLASP-expressing clones can be used to test for homotypic and/or heterotypic calcium dependent cell adhesion using the "cell aggregation assay" described for cadherin molecules (Murphy-Erdosh, C. *et al.*, 1995, J. Cell Biol. 129: 1379-1390).

Several approaches can be used to identify the amino acids involved in the binding domains. Soluble fusion molecules (*e.g.*, EC12-IgG, ECC-IgG, ECM-IgG, and GST-EC12), peptides, and peptide-specific anti-CLASP antibodies are available for blocking experiments in the above-described assay. Transfectants generated by site-directed mutagenesis can also be used.

(C) Membrane Anchoring/Cytoskeletal Interactions

Interestingly, tyrosine-phosphorylated ITAMs interact with actin cytoskeleton upon activation of mature T lymphocytes (Rozdzial, M. M., 1995, Immunity 3: 623-633). Since human CLASPs contain both ITAMs and coiled-coil domains which have been shown to interact with cytoskeletal proteins, CLASPs are believed to play an important role in modulating cell surface molecule expression by re-organizing cytoskeletal structure.

F-actin microfilament cytoskeletal organization has been known to be involved in the modulation of cell surface molecule expression. WASP, a GTPase-binding protein, plays a critical role in the formation of actin microspikes in response to external stimuli and ectopic expression of WASP induces the formation of F-actin filament clusters that overlap with the expressed WASP itself. Another WASP family protein, N-WASP, has also been shown to play important roles in filopodium formation. Both of these proteins cause actin polymerization, but with different features when they are expressed in cells; WASP mainly localizes at perinuclear areas and causes actin clustering, but most N-WASP is present at plasma membranes and induces filopodium formation (Miki, H.; 1998, Nature 391: 93-6). Both WASP and N-WASP, contain a proline-rich domain which could interact with the SH3 domain present in all the human CLASPs. CLASP-2 may interact with F-actin filament through CLASP-2 binding to WASP or WASP-like proteins.

Standard assays can be used for detecting CLASP protein interaction with cytoskeletal proteins. These assays include co-sedimentation assays, far western blot analysis (Ohba, T., 1998, Anal. Biochem. 262: 185-192), surface pasman resonance, F-actin staining

with phalloidin in CLASP-transfected lymphocytes (*e.g.*, Small, J. *et al.* 1999, Microsc. Res. Tech. 4: 3-17), and immunocytal analysis of subcellular distribution of focal adhesion proteins (such as paxillin, tensin, vinculin, talin, and FAK in CLASP-transfected lymphocytes; see, *e.g.*, Ridyard, M.S., 1998, Biochem. Cell Biol. 76: 45-58).

5.2.2. CLASP-2 Exon Structure and Genomic Domains

Alternative splicing is likely to represent a regulatory switch that governs different functions of CLASP-2 in immune responses Additionally, alternative splice variants affecting the untranslated regions of an RNA can be a way of regulating RNA stability.

As noted *supra*, CLASP-2 gene expression is characterized by alternative exon usage. Intron/exon structure can be predicted by computer analysis of genomic DNA, however, splice junctions and alternative splicing can only be elucidated by comparison of genomic clones to cDNA clones. Alternative splicing and RNA editing are mechanisms generate a variety of proteins from the same gene. An example for how alternative splicing is used to generate thousands of different proteins from only a few genes is represented by the Neurexin gene family (for review of Neurexins, see Missler M. and Suedhof, T., 1998, Trends in Genetics, 14: 20-25). Comparative analysis of CLASP-2 genomic clones and cDNA clones revealed that CLASP-2 is composed of numerous exons and that distinct CLASP-2 transcripts are generated by alternative splicing. The protein encoding portion of CLASP-2 is covered by at least 14 exons (FIG. 6A).

Numerous diseases are caused or are thought to be caused by splice site mutations that can cause exon skipping or otherwise result in a truncated protein product Some of these diseases include, e.g., Marfan Syndrome (Liu W, et al., 1997, Nat. Genet. 16: 328-9), Hunter disease (Bonucelli G, et al., 2000, Hum. Mutat. (Online) 2000 15(4): 389, Duchenne muscular dystrophy (Wibawa T, et al., 2000, Brain Dev. 22(2): 107-112), Myelomonocytic leukemia (Wutz D, et al., 1999, Leuk. Lymphoma 35: 491-9.), and Isovaleric acidemia (Vockley J, et al., 2000, Am. J. Hum. Genet. 66: 356-67). This is especially true for genes composed of many exons (such as CLASP-2). The genomic sequence around CLASP-2 exon/intron boundaries is useful for diagnostic approaches towards the identification of diseases caused by splice site mutations. The abundance or presence of CLASP-2 isoforms in cell populations (e.g., hematopoietic cells, lymphocytes) is correlated with a disease state by comparing the abundance of CLASP-2 in cells from

subjects suffering from the disease with the level of CLASP-2 in cells from healthy subjects. This can be accomplished by utilizing any number of assays (e.g., PCR).

Alignment of the CLASP-2 intron/exon splice sites with the CLASP-2 protein sequence and the finding of conserved exon/intron boundaries within the CLASP gene family (FIG. 6) suggest that specific CLASP-2 exons encode functionally distinct protein domains (see FIG. 6 and Example 4). ITAM and DOCK motifs 1 and 2 are encompassed by splice sites (amino acid residues 946 and 1063); DOCK motif 3 and COILED-COIL motif 1 and 2 are also encompassed by splice sites (amino acid residues 1102, 1170 and 1246, respectively).

CLASP-2 alternative transcripts are summarized in FIG. 3 and FIG.11B. Briefly, one alternative exon missing in CLASP-2A is present in CLASP-2B and CLASP-2D. This exon contains the DNA portion encoding the ITAM motif and DOCK motif 1. The CLASP-2D protein product does not contain the C-terminal 38 amino acids of CLASP-2A and CLASP-2B: Thus, a PDZ binding motif (SSVV; amino acid residue 1286 through 1289) that is only present in the CLASP-2A/B-specific C-terminal end is missing in the CLASP-2D gene product. The presence or absence of this PDZ binding motif can be attributed to alternative RNA processing. Additionally, a CLASP-2 alternative transcript has been found that deletes nucleotides 209-291 that results in a premature stop codon. The protein encoded by this transcript appears to be a soluble form of CLASP-2 that may regulate (*e.g.*, is an antagonist or an agonist) the function other CLASP family members and isoforms.

5.2.3. CLASP Superfamily Members

As is illustrated in FIG. 5, CLASP-2 is a member of a superfamily of immune-cell associated proteins with similar motifs. CLASP-1 was described in U.S.S.N. 09/411,328, filed October 1, 1999. CLASP-1 uniquely among the known CLASPs contains SH3 binding domain motifs. CLASP-2A, -B, -C, and -E polypeptides have no adaptor binding sites or SH3 binding domains found in CLASP-1. CLASP-3, CLASP-4, CLASP-5 and CLASP-7 are described in copending U.S.S.N. 60/182,296, filed February 14, 2000, and which is incorporated by reference herein in its entirely for all purposes.

5.3. CLASP-2 mRNA Expression

As described in Example 2, CLASP-2 mRNA expression was assayed in tissues and cell lines by Northern analysis. The results are shown in FIG. 4A and B. The

results of Northern Analysis of CLASP-2 expression and expression of other members of the CLASP family are summarized in Table 2.

Table 2

Tissue/Cell Line[1]	CLASP					
	1	2[3,4]	3	4	5	7
PBL	+[2]	-	-	+++	++	-
Lung	-	+	-	-	-/+	+++
Placenta	-/+	+++	+	-/+	+	+
Sm Intestine	-/+	-	-	-	-/+	+
Liver	-/+	-/+	-/+	-	-/+	+
Kidney	-/+	+	+++	-/+	+	++
Spleen	++	-	-	-/+	+	-/+
Thymus	++	-	-	-/+	+	-
Colon	-	-	-	-	-	-
Skel Muscle	-	-/+	++	-	-	-/+
Heart	-/+	++	+++	-/+	-	+++
Brain	+++	-/+	-/+	-	-	-
Jurkat	++	++	++	+	-	-
MV411	++	-	++	+	+	+
THP1	++	-	-	-	-	-/+
HL60	-	-	-	-	-/+	-
9D10	++	++[5]	+	+	+	+
3A9	+	-/+	-	-	-	-
CH27	+	-/+	-	-	-	-
293	-	++	+++	+	-	+

1. Jurkat = human T cell line; MV4-11 = B myelomonocyte; 9D10 = B cell line; THP-1 = monocyte; 3A9 = mouse T cell; CH27 = mouse B cell line; HL60 = human promyelocyte; 293 = embryonic kidney epithelial cells (293)

2. Table Legend (based on Northern blot results): - = no expression; -/+ = low expression; + =medium expression; ++ medium high expression; +++ high expression.

3. A CLASP-2 EST (EST 815795) was identified from a bone marrow cDNA library.

4. The probe used (HC2.2) did not distinguish between CLASP-2A, -2B, -2C and 2D.. This probe encompasses nucleotides 3920 to 4650 (731 bp long) from CLASP-2A cDNA.

5. In RNA from 9D10, the major transcript runs substantially shorter than the major transcripts seen in Jurkat and 293 cells; however, the longer transcript is also present in 9D10. Hybridization of probe HC2.2 with 9D10 total RNA reveals at least 3 different transcripts. See FIG. 4B

As indicated in Table 2 and shown in FIG. 4, CLASP-2 is expressed most strongly in placenta followed by lung, kidney and heart; CLASP-3 is expressed strongly in kidney and heart, and less strongly in placenta and skeletal muscle ; CLASP-4 is expressed exclusively in peripheral blood lymphocytes; CLASP-5 is expressed strongly in peripheral blood leukocytes, present in placenta, kidney, spleen and thymus, and weakly in lung, small

intestine and liver. It is not expressed in brain, heart, skeletal muscle and large intestine; CLASP-7 is expressed strongly in lung, heart, liver and kidney, but not in PBL, brain or thymus.

Differences in tissue expression patterns for different CLASP proteins indicate different CLASPs have differential roles in immune function and, accordingly, can be separately targeted to achieve different functions. For example, since CLASP proteins are necessary for proper function or signaling by the T cell receptor (TCR), the tissue specific distribution of different CLASPs permits differential modulation of the immune response in different tissues. Since CLASP-2 is present in heart, blocking CLASP-2 function or expression is useful to selectively block immune response in the heart (for example, to selectively stop immune response in the heart compartment, *e.g.*, following cardiac transplant rejection or post-MI inflammation, without compromising immunity elsewhere. Similarly, blocking CLASP-3 can block rejection of the kidney following kidney transplant. Furthermore, by adjusting the level of inhibition, the degree of immune blockage versus response can be modulated in the compartments represented by each CLASP.

5.4. CLASP-2 Polynucleotides And Methods Of Use

The present invention provides a variety of CLASP-2 polynucleotides and methods for using them. In one aspect, the polynucleotide of the invention encodes a polypeptide comprising at least a fragment (*e.g.*, an immunogenic fragment) of a CLASP-2 protein (*e.g.*, at least a fragment of SEQ. ID. NO: 2, 4, 6 or 10) or variant thereof. In another aspect, the molecules that comprise a CLASP-2 polynucleotide that, while not necessarily encoding a CLASP-2 protein or fragment, is useful as a probe or primer for detecting CLASP-2 expression, for inhibition of CLASP-2 expression (*e.g.*, antisense or ribozyme-mediated inhibition), for gene knockout, and the like.

5.4.1. CLASP-2 Polynucleotides

The invention also provides isolated or purified nucleic acids having at least 8 nucleotides (*i.e.*, a hybridizable portion) of a CLASP-2 sequence or its complement; in other embodiments, the nucleic acids consist of at least about 25 (continuous) nucleotides, about 50 nucleotides, about 100 nucleotides, about 150 nucleotides, about 200 nucleotides, about 250 nucleotides, about 500 nucleotides, about 550 nucleotides, about 600 nucleotides, or about 650 nucleotides or more of a CLASP-2 sequence, or a full-length CLASP-2 coding sequence. In another embodiment, the nucleic acids are smaller than about 35, about 200 or about 500

nucleotides in length. Polynucleotides can be single or double stranded, and may be DNA, RNA, PNA or a hybrid molecule.

In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least about 10, 25, 50, 100, 150, 200, 250, 500, 550, 600, or 650 nucleotides or the entire coding region of a CLASP-2 coding sequence. Usually, the isolated polynucleotide is less than about 100 kbp, generally less than about 50 kbp, and often less than about 20 kbp, less than about 10 kbp, less than about 5 kbp, or less than about 1000 nucleotides in length.

In a specific embodiment, a nucleic acid that is hybridizable to a CLASP-2 nucleic acid or its complement, or to a nucleic acid encoding a CLASP-2 derivative, under conditions of low stringency is provided. Derivatives of CLASP-2 contemplated include, but are not limited to, splice variants of a gene encoding a CLASP-2, other members of a CLASP-2 gene family which differ from one of the CLASP-2 nucleotide or amino acid sequences disclosed herein by the insertion or deletion of one or several domains, and the like.

In one embodiment, the CLASP-2 polynucleotide is identical or exactly complementary to SEQ. ID NO: 1, 3, 5 or 9 or selectively hybridizes to an aforementioned sequence. In various embodiments, the polynucleotide is identical or exactly complementary to, or selectively hybridizes to, the nucleotide sequence encoding a particular protein domain or region, or a particular gene exon of the CLASP-2 mRNA or genomic sequence. Such polynucleotides are particularly useful as probes, because they can be selected to identify a defined species of CLASP-2.

In addition to the polypeptide and polynucleotide sequences specifically exemplified herein, the invention contemplates CLASP-2 homologues from other species, allelic and splice variants, and other variants disclosed herein. The CLASP-2 gene exhibits evidence of alternative splicing of transcripts.

For example, CLASP-2A and CLASP-2C are related to each other as apparent splice variants, with CLASP-2C containing an exon not found in CLASP-2A. The exon sequence is 5'-AGG GAT TTT GAG AGG CTG GCC CAT CTG TAT GAC ACG CTG CAC CGG GCC TAC AGC AAA GTG ACC GAG GTC ATG CAC TCG GGC CGC AGT TNC TGG GGA CCT ACT TCC GGG TAG CCT TCT TCG GGC AG-3' (encoding the peptide sequence: RDFERLAHLYDTLHRAYSKVTEVMHSGRRLLGTYFRVAFFGQGF). It will be apparent to one of skill that, by using polynucleotide probes or primers corresponding to the nucleic acid sequence above, or by using antibodies that specifically

recognize the peptide above, or those polynucleotide probes or primers shown in Table 3 below, it is possible to distinguish between different CLASP isoforms(*e.g.*, to detect differential expression).

Table 3

	Found in/will detect	Exemplary Probe/Primer (5' – 3')	Notes/Comments
1	full length hC2A	F1: CCCAGATTTTTATGATGAG R1: GATAATGACAAAGTTCTGAC	
2	full length hC2D	F2: CTGGAAATCTTGACAAAAATGC R2: GTCTTTTTAATACAGATGTGG	
3	hC2B, hC2C, hC2E	F3: GAGAGGCTGGCCCATCTGTATG R3: ATCTTCAAAGAATCCCTGCC	Distinction based upon product size differences following PCR
4	hC2D	F4: GAAGCAGTCCAGTGGGAGCCG R4: GCCTCCCCGGCTCCTCCTCAGG	Recognizes hC2D-specific insertion
5	hC2D	F3: GAGAGGCTGGCCCATCTGTATG R5: CCTCCACATCTGTTTCACTGTC	
6	hC2E	F5: CTCCATGATGGAAGACGTGGG R6: GATGAGCTCGTAGCGCTCGGC	Spans deletion unique to hC2E. Distinction based upon product size differences following PCR
7	hC2B	F6: CATTGGCGTTTAAGCTCCTG R3: ATCTTCAAAGAATCCCTGCC	F6 primer spans deletion unique to hC2E
8	hC2A	F7: GGACCCATAGTTCATGATCG R4: CTTCATCTTCAAGAAATCCCTC	R4 primer spans the region where other CLASPs have an insert

5

5.4.1.1. Substantial Identity

In some embodiments, the CLASP-2 polynucleotides of the invention are substantially identical to SEQ ID NOs: 1, 3, 5, or 9, or to a fragment thereof.

An indication that two nucleic acid sequences are substantially identical is that the two polynucleotides have a specified percentage sequence identity *e.g.*, usually at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 98 identity over a specified region when optimally aligned.

Another indication that two nucleic acid sequences are substantially identical is that a polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication

that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

Yet another indication that two nucleic acid sequences are substantially identical (*e.g.*, a naturally occurring allele of the CLASP-2 sequence of SEQ ID NO: 1) is that the same primers can be used to amplify the sequence. For example, CLASP-2 polynucleotides can be PCR amplified from cDNA derived from human lymphocytes using the primer pairs shown in Table 3.

The primers of Table 3 are also useful for amplification of CLASP-2 splice variants. Another indication that two nucleic acid sequences are substantially identical is that they selective hybridize under stringent conditions (*i.e.*, one sequence hybridizes to the complement of the second sequence), as described *infra*.

5.4.1.2. Selective Hybridization

The invention also relates to nucleic acids that selectively hybridize to exemplified CLASP-2 sequences (including hybridizing to the exact complements of these sequences). Selective hybridization can occur under conditions of high stringency (also called "stringent hybridization conditions"), moderate stringency, or low stringency.

5.4.1.2.1. High Stringency

"Stringent hybridization conditions" are conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but not to other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides)

and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For high stringency hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary high stringency or stringent hybridization conditions include: 50% formamide, 5x SSC and 1% SDS incubated at 42° C or 5x SSC and 1% SDS incubated at 65° C, with a wash in 0.2x SSC and 0.1% SDS at 65° C. In a specific embodiment, a nucleic acid which is hybridizable to a CLASP-2 nucleic acid under the following conditions of high stringency is provided: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 8-16 h at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 65°C for 15-30 h in a solution containing 2X SSC, 0.1% SDS. This is followed by a wash in 0.2X SSC and 0.1% at 50°C for 15-30 min before autoradiography.

5.4.1.2.2. Moderate Stringency

In another specific embodiment, a nucleic acid, which is hybridizable to a CLASP-2 nucleic acid under conditions of moderate stringency is provided. Examples of procedures using such conditions of moderate stringency are as follows: Filters containing DNA are pretreated for 6 h at 55°C in a solution containing 6X SSC, 5X Denhart's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 12-16 h at 55°C, and then washed twice for 30 minutes at 50°C in a solution containing 1X SSC and 0.1% SDS. Filters are blotted dry and exposed for autoradiography. Other conditions of moderate stringency which can be used are well-known in the art. Washing of filters is done at 45°C for 1 h in a solution containing 0.2X SSC and 0.1% SDS.

5.4.1.2.3. Low Stringency

By way of example and not limitation, procedures using such conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. U.S.A. 78: 6789-6792): Filters containing DNA are pretreated for 6 h at 40 C in a solution

containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 g/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40 C, and then washed for 1.5 h at 55 C in a solution containing 2X SSC and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 30 minutes at 50-55°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 60-65°C and reexposed to film. Other conditions of low stringency that can be used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

5.4.1.3. CLASP-2 Variants and Fragments

The CLASP-2 variants of the invention can contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. CLASP-2 polynucleotide variants can be produced for a variety of reasons, *e.g.*, to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Exemplary CLASP-2 polynucleotide fragments are preferably at least about 15 nucleotides, and more preferably at least about 20 nucleotides, still more preferably at least about 30 nucleotides, and even more preferably, at least about 40 nucleotides in length, or larger 50, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 nucleotides. In one embodiment, exemplary fragments include fragments having at least a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600 to the end of SEQ ID NO: 1 or SEQ ID NO: ____or comprising the cDNA coding sequence in the deposited clones. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In other embodiments, CLASP-2 polynucleotides of the invention are other than SEQ ID NO:1 or fragments of SEQ ID NO:1.

As shown in FIG 11 above, there are at least three CLASP-2 full length cDNA isoforms (A + Z, B + Z, and C + Z). Each of the isoforms uses a unique first exon (labelled exon 1A, 1B, and 1C) (*see* FIG. 11 and Table 4 below).

Table 4: CLASP-2 Isoforms

CLASP-2 Isoform	FIG 11C Schematic	Nucleotides
Isoform 1	A + Z	–182 to 6690
Isoform 2	B + Z	–219 to 6690
Isoform 3	C + Z	–143 to 6690

In one embodiment, the CLASP-2 polynucleotide has the sequence shown in FIG. 11 (Isoform 1, Isoform 2, or Isoform 3 as indicated in Table 4 above) or a fragment of the sequence shown in FIG. 11 comprising at least about 1, 5, 10, 25 or 50 or more continguous nucleotides from nucleotides –182 to 1883 of Isoform 1, nucleotides –219 to 1883 of Isoform 2, or nucleotides –143 to 1883 of Isoform 3.

In another embodiment, CLASP-2 primers or probes comprise at least about 5, 10, 25 or 50 or more continguous nucleotides from nucleotides –182 to 1883 of Isoform 1, nucleotides –219 to 1883 of Isoform 2, or nucleotides –143 to 1883 of Isoform 3 as shown in FIG. 11 and Table 4 above alone or in combination with SEQ ID NO:1 or a fragment of SEQ ID NO:1.

In an aspect, the invention provides antibodies or binding fragments that bind the CLASP-2 isoforms 1-3. In another embodiment, the invention provides antibodies that specifically bind to the CLASP-2 isoforms shown in FIG. 11 but not to the polypeptide encoded by SEQ ID NO:1

In one embodiment, the CLASP-2 variants differ from those shown in FIG. 1 or FIG. 11 (SEQ ID NOs 1, 3, 5, 7 9, _____) by virtue of incorporating a different combination of exons than found in the exemplified sequences. For example, 81g01 (Genbank Accession Number AF85864; Locus HUMYN81g01; 526 bp; EST sequence submitted August 29, 1998 by Washington University at St. Louis; see FIG. 3A and FIG. 3B) is a variant of hCLASP-2 on the basis of CLASP-2 identity along certain stretches of the sequence. From 5' to 3', it begins with a 315 nucleotide stretch which is identical to CLASP-2A. It then has a gap relative to CLASP-2A that is identical to the GAP in another CLASP-2 isoform, hCLASP-2D (KIAA1058). In place of that gap, a 16 amino acid insert (48 nucleotides) is present which is not found in other isoforms, followed by another

approximately 150 bp stretch of nucleotides once again identical to CLASP-2A. This is characteristic of an alternate splice due to the specific sequence identity on both sides of a differential stretch of nucleotides.

Using known methods of protein engineering and recombinant DNA technology, variants can be generated to improve or alter the characteristics of the CLASP-2 polypeptides. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the CLASP-2 protein without substantial loss of biological function.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities can still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking Nor C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes CLASP-2 polypeptide variants which show biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. *et al.*, Science 247: 1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at 30 specific positions of a cloned gene to identify regions critical for protein function. For example., site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells,

1989, Science 244: 1081-1085) The resulting mutant molecules can then be tested for biological activity.

In various embodiments, CLASP-2 polynucleotide fragments include coding regions for, or regions hybridizable to, the CLASP-2 structural or functional domains described *supra*. As set out in the Figures, such preferred regions include the following domains/motifs: ITAM, DOCK, COILED/COILED, and PBM. Thus, for example, polypeptide fragments of CLASP-2 as shown in FIG. 1 and FIG. 11-(SEQ ID NO: 2, 4, 6 10, _____) falling within conserved domains are specifically contemplated by the present invention (see FIG. 3). Moreover, polynucleotide fragments encoding these domains are also contemplated. Such polypeptide fragments find use, for example, as inhibitors of CLASP-2 function in CLASP-2-expressing cells.

5.4.2. Uses of CLASP-2 Polynucleotides

The CLASP-2 polynucleotides of the invention are useful in a variety of applications. In one aspect of the invention, the polypeptide-encoding CLASP-2 polynucleotides of the invention are used to express CLASP-2 polypeptides (*e.g.*, as described herein) for example to produce anti-CLASP-antibodies or for use as therapeutic polypeptides. In another aspect, the CLASP-2 polynucleotide or fragments thereof can be used for diagnostic purposes (*e.g.*, as probes for CLASP-2 expression). In particular, since CLASP-2s can be expressed in lymphocytes, a CLASP-2 polynucleotide can be used to detect the expression of CLASP-2 as a lymphocyte marker. For diagnostic purposes, a CLASP-2 polynucleotide can be used to detect CLASP-2 gene expression or aberrant CLASP-2 gene expression in disease states. In another aspect, the CLASP-2 polynucleotide or fragments are used for therapeutic purposes. For example, included in the scope of the invention are methods for inhibiting CLASP-2 expression, *e.g.*, using oligonucleotide sequences, such as antisense RNA and DNA molecules and ribozymes, that function to inhibit expression of CLASP-2. In another aspect, CLASP-2 polynucleotides can be used to construct transgenic and knockout animals, *e.g.*, for screening of CLASP-2 agonists and antagonists. In another aspect, CLASP-2 polynucleotides can be used for screening of CLASP-2 agonists and antagonists.

5.4.2.1. Use of CLASP-2 P lynucleotides for Detection, Diagnosis, and Treatment

The CLASP-2 polynucleotides of the invention are useful for detection of CLASP-2 expression in cells and in the diagnosis of diseases or disorders (*e.g.*,

immunodeficient states) resulting from aberrant expression of CLASP-2. Aberrant expression of CLASP-2 mRNA or protein means expression in lymphocytes (*e.g.*, T lymphocytes or B lymphocytes) or other CLASP-2 expressing cells of at least 2-fold, preferably at least 5-fold greater or less than expression in control lymphocytes obtained from a healthy subject. CLASP-2 polypeptide expression is easily measured by ELISA using anti-CLASP-2 antibodies of the invention. CLASP-2 mRNA expression (including expression of specific species or splice variants of CLASP-2) can be measured by quantitative Northern analysis or quantitative PCR, LCR, or other methods, using the probes and primers of the invention.

In one embodiment, the assays of the present invention are amplification-based assays for detection of an CLASP-2 gene product. In an amplification based assay, all or part of a CLASP-2 mRNA or cDNA (hereinafter also referred to as "target") is amplified, and the amplification product is then detected directly or indirectly. When there is no underlying gene product to act as a template, no amplification product is produced (*e.g.*, of the expected size), or amplification is non-specific and typically there is no single amplification product. In contrast, when the underlying gene or gene product is present, the target sequence is amplified, providing an indication of the presence and/or quantity of the underlying gene or mRNA. Target amplification-based assays are well known to those of skill in the art.

The present invention provides a wide variety of primers and probes for detecting CLASP-2 genes and gene products. Such primers and probes are sufficiently complementary to the CLASP-2 gene or gene product to hybridize to the target nucleic acid. Primers are typically at least 6 bases in length, usually between about 10 and about 100 bases, typically between about 12 and about 50 bases, and often between about 14 and about 25 bases in length, often PCR primers of 15-30 (*e.g.*, 18-22 nucleotides) are used. However, the length of primers can be adjusted by one skilled in the art. One of skill, having reviewed the present disclosure, will be able, using routine methods, to select primers to amplify all, or any portion, of the CLASP-2 gene or gene product, or to distinguish between variant gene products, CLASP-2 alleles, and the like. Single oligomers (*e.g.*, U.S. Pat. No. 5,545,522), nested sets of oligomers, or even a degenerate pool of oligomers can be employed for amplification.

It will be appreciated that probes and primers can be selected to distinguish between species and splice variants based on the guidance of this disclosure, by targeting

primers or probes to differentially used exons (or exon-exon junctions that differ between variants).

Methods can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to an CLASP-2 gene under conditions such that hybridization and amplification of the CLASP-2-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. See U.S. Pat. Nos. 4,683,195 and 4,683,202, Landegran *et al.*, 1988, Science 241: 1077-1080; Nakazawa *et al.*, 1994, Proc. Natl. Acad. Sci. U.S.A. 91: 360-364, Abravaya *et al.*, 1995, Nucleic Acids Res. 23: 675-682).

Because CLASP-2 gene products are expressed in the immune system (*e.g.*, T lymphocytes, B lymphocytes and macrophages), expression will be typically assayed in these cells. Methods which are well known to those skilled in the art can be used to isolate lymphocytes, macrophages, and alike (*See*, *e.g.*, Coligan, J. E., *et al.* (eds.), 1991, Current Protocols in Immunology, John Wiley & Sons, NY; this reference is incorporated by reference for all purposes). In one embodiment, assays are carried out on biopsy or autopsy-derived tissue.

In various embodiments, CLASP-2 gene expression is detected by hybridization of a detectable probe to mRNA or cDNA obtained from cells (*e.g.*, lymphocytes). A variety of methods for specific DNA and RNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al.*, *supra*). Hybridization based assays refer to assays in which a probe nucleic acid is hybridized to a target nucleic acid, forming a hybridization complex. Usually the nucleic acid hybridization probes of the invention are entirely or substantially identical to a contiguous sequence of the CLASP-2 gene or RNA sequence. Preferably, nucleic acid probes are at least about 50 bases, often at least about 20 bases, and sometimes at least about 200 bases, at least about 300-500 nucleotides or more in length. Various hybridization techniques are well known in the art, and are in fact the basis of many commercially available diagnostic kits.

Methods of selecting nucleic acid probe sequences for use in nucleic acid hybridization are discussed in Sambrook *et al.*, *supra*. In some formats, at least one of the target and probe is immobilized. The immobilized nucleic acid can be DNA, RNA, or another oligo- or poly-nucleotide, and can comprise natural or non-naturally occurring

nucleotides, nucleotide analogs, or backbones. Such assays can be in any of several formats including: Southern, Northern, dot and slot blots, high-density polynucleotide or oligonucleotide arrays (*e.g.*, GeneChips™ Affymetrix), dip sticks, pins, chips, or beads. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits. Hybridization techniques are generally described in Hames *et al.*, ed., 1985, Nucleic Acid Hybridization, A Practical Approach IRL Press; Gall and Pardue, 1969, Proc. Natl. Acad. Sci. U.S.A., 63: 378-383; and John *et al.*, 1969, Nature, 223: 582-587.

A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, one common format is direct hybridization, in which a target nucleic acid is hybridized to a labeled, complementary probe. Typically, labeled nucleic acids are used for hybridization, with the label providing the detectable signal. One method for evaluating the presence, absence, or quantity of CLASP-2 mRNA is carrying out a Northern transfer of RNA from a sample and hybridization of a labeled CLASP-2 specific nucleic acid probe. A useful method for evaluating the presence, absence, or quantity of DNA encoding CLASP-2 proteins in a sample involves a Southern transfer of DNA from a sample and hybridization of a labeled CLASP-2 specific nucleic acid probe.

Other common hybridization formats include sandwich assays and competition or displacement assays. Sandwich assays are commercially useful hybridization assays for detecting or isolating nucleic acid sequences. Such assays utilize a "capture" nucleic acid covalently immobilized to a solid support and a labeled "signal" nucleic acid in solution. The biological or clinical sample will provide the target nucleic acid. The "capture" nucleic acid and "signal" nucleic acid probe hybridize with the target nucleic acid to form a "sandwich" hybridization complex. To be effective, the signal nucleic acid cannot hybridize with the capture nucleic acid.

In one embodiment, CLASP-2 polypeptides or polynucleotides are useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the activation, differentiation of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders can be genetic, somatic, such as cancer or some autoimmune disorders, acquired (*e.g.*, by chemotherapy or toxins), or infectious.

In another embodiment, CLASP-2 polynucleotides or polypeptides are useful in treating or detecting deficiencies or disorders of hematopoietic cells. CLASP-2

polypeptides or polynucleotides could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (*e.g.*, agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

In one embodiment, CLASP-2 polynucleotides or polypeptides are useful in treating or detecting autoimmune diseases. The term "autoimmune disease" as used herein has the normal meaning in the art and refers to a spontaneous or induced malfunction of the immune system of mammals in which the immune system fails to distinguish between foreign immunogenic substances within the mammal and/or autologous ("self") substances and, as a result, treats autologous ("self") tissues and substances as if they were foreign and mounts an immune response against them. Autoimmune disease is characterized by production of either antibodies that react with self tissue, and/or the activation of immune effector T cells that are autoreactive to endogenous self antigens. Three main immunopathologic mechanisms act to mediate autoimmune diseases: 1) autoantibodies are directed against functional cellular receptors or other cell surface molecules, and either stimulate or inhibit specialized cellular function with or without destruction of cells or tissues; 2) autoantigen--autoantibody immune complexes form in intercellular fluids or in the general circulation and ultimately mediate tissue damage; and 3) lymphocytes produce tissue lesions by release of cytokines or by attracting other destructive inflammatory cell types to the lesions. These inflammatory cells in turn lead to production of lipid mediators and cytokines with associated inflammatory disease.

Since many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of CLASP-2 polypeptides or polynucleotides that can inhibit an immune response, particularly the proliferation, or differentiation of T-cells, can be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by CLASP-2 include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid

syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, can also be treated by CLASP-2 polypeptides or polynucleotides. Moreover, CLASP-2 can be used to treat anaphylaxis or hypersensitivity to an antigenic molecules.

In one embodiment CLASP-2 polynucleotides or polypeptides are used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of CLASP-2 polypeptides or polynucleotides that inhibits an immune response, particularly the proliferation, differentiation of T-cells, can be an effective therapy in preventing organ rejection or GVHD.

Similarly, in another embodiment, CLASP-2 polypeptides or polynucleotides are used to modulate inflammation. The term "inflammation" refers to both acute responses (*i.e.*, responses in which the inflammatory processes are active) and chronic responses (*i.e.*, responses marked by slow progression and formation of new connective tissue). Acute and chronic inflammation can be distinguished by the cell types involved. Acute inflammation often involves polymorphonuclear neutrophils; whereas chronic inflammation is normally characterized by a lymphohistiocytic and/or granulomatous response. Inflammation includes reactions of both the specific and non-specific defense systems. A specific defense system reaction is a specific immune system reaction response to an antigen (possibly including an autoantigen). A non-specific defense system reaction is an inflammatory response mediated by leukocytes incapable of immunological memory. Such cells include granulocytes, macrophages, neutrophils and eosinophils.

For example, CLASP-2 polypeptides or polynucleotides can inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (*e.g.*, septic shock, sepsis, or systemic

inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (*e.g.*, TNF or IL-1.). Examples of specific types of inflammation are diffuse inflammation, focal inflammation, croupous inflammation, interstitial inflammation, obliterative inflammation, parenchymatous inflammation, reactive inflammation, specific inflammation, toxic inflammation and traumatic inflammation.

In another embodiment CLASP-2 polypeptides or polynucleotides are used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases can be treated. The immune response can be increased by either enhancing an existing immune response, or by initiating a new immune response. CLASP-2 polypeptides or polynucleotides can be used to treat or detect any of these symptoms or diseases.

5.4.2.2. Use of CLASP-2 Polynucleotides in Screening

The presence or absence of hCLASP-2 nucleotide and amino acid sequences in a biological sample can be used in screening assays as medical diagnostics to aid in clinical decision-making. In one embodiment, hCLASP-2-based diagnostics involves screening assays for vaginal bleeding of unknown cause. In several examples discussed below, the cause of the bleeding can be in part differentiated by knowledge of whether the vaginal bleeding contains placental components (Hart FD, Ed., 1985, French's Index of Differential Diagnosis, 12th Ed. John Wright & Sons, pp. 561-63). In these cases, the high expression of hCLASP-2 nucleotide sequences in placenta relative to its low expression in blood (FIG. 4A) will allow the detection of the presence of placenta based on the presence of the hCLASP-2 nucleotide or protein. Such detection can be achieved by quantitative RT-PCR, Northern analysis, Western analysis, ELISAs, and fluorescence activated cell sorting (FACS) by using labeled anti-hCLASP-2 antibodies (Sambrook *et al.*, 1989, Molecular Cloning, 2nd Ed., Cold Spring Harbor Lab. Press; Harlow et. al., 1988, Antibodies, a laboratory manual, Cold Spring Harbor Lab. Press).

For example, hCLASP-2 can be used in the following screening assays:

(1) A woman gives birth and presents with post-partum bleeding. In this case the presence of placental tissue indicates a condition called "retained products of

conception" that requires surgical evacuation of the uterus (Decherney and Pernol, Eds., 1996, Current Obstetric & Gynecologic Diagnosis & Treatment, 8th Ed. McGraw Hill).

(2) A pregnant woman suffers from vaginal bleeding of unknown origin. In this case the presence of placental tissue indicates a condition called "threatened abortion" that implies a poor prognosis for carrying the fetus to term (Decherney and Pernol, Eds., 1996, Current Obstetric & Gynecologic Diagnosis & Treatment, 8th Ed. McGraw Hill).

(3) A woman of child bearing age presents with vaginal bleeding and is found to have a positive pregnancy test without evidence of an intra-uterine pregnancy. In this case, the most serious of the differential diagnoses is ectopic pregnancy, a medical emergency. However, another common diagnosis is a completed abortion or miscarriage. The presence of products of conception (*i.e.* placenta) in the vaginal bleeding strongly favors the diagnosis of completed abortion over that of ectopic pregnancy (Decherney and Pernol, Eds., 1996, Current Obstetric & Gynecologic Diagnosis & Treatment, 8th Ed. McGraw Hill).

In another embodiment, hCLASP-2-based diagnostics involve screening assays to determine injury to vital tissues that express hCLASP-2 at high levels. Such tissues include kidney, heart, and lung (Fig 4A). Injury to these tissues can result in leakage of cells and cellular constituents including hCLASP-2 into surrounding fluids (specified below). Detection of abnormally high levels of hCLASP-2 protein in these surrounding fluids by Western analysis or ELISA, or detection of abnormally high levels of hCLASP-2 RNA in these fluids by RT-PCR or Northern analysis is expected to aid in the diagnosis of tissue injury.

In the case of renal injury, the hCLASP-2 nucleotide or amino acid sequences or fragments thereof would be expected to appear in the urine. Detection of abnormally high levels of hCLASP-2 can aid in the diagnosis of both nephritis and tubular necrosis, and differentiate from non-renal causes of proteinuria. Early diagnosis of nephritis is of particular value in patients with clinical signs and symptoms suggestive of systemic lupus erythematosis in whom early diagnosis and treatment of lupus nephritis can prevent irreversible kidney damage (Cameron J.S., 1999, J Nephrol 12 Suppl 2: S29-41). While tubular necrosis currently cannot be reversed by pharmacotherapy, differentiation of tubular necrosis from pre-renal failure is critical in formulating a treatment plan for oligouric hospitalized patients (Bidani A. and Churchill P.C., 1989, Dis Mon 35: 57-132).

In the case of myocardial injury, the hCLASP-2 nucleic or amino acid sequence or fragments thereof are expected to appear in the blood. This is analogous to current standard practice of monitoring for other elevated levels myocardial proteins (*e.g.*,

creatine kinase, troponin) in the blood following myocardial infarction and ischemia by standard ELISA or electrophoretic methodologies (Fauci *et al.*, (eds.), 1998, Harrison's Principles of Internal Medicine, 14th Ed., McGraw Hill, pp. 1352-1375). The presence of hCLASP-2 in cardiac muscle and its absence in skeletal muscle and blood makes hCLASP-2 an ideal marker to diagnose and monitor myocardial injury.

Unlike myocardial injury, pulmonary injury is not routinely diagnosed by assaying serum for lung-specific proteins. By analogy to myocardial infarction, pulmonary infarction also releases lung-specific proteins and cells into systemic circulation. Pulmonary infarction due to pulmonary embolism (PE) or pneumonia is expected to release hCLASP-2-bearing cells or protein/peptides into systemic circulation. Detection of hCLASP-2 protein in serum or RNA in blood can aid in the diagnosis of pulmonary infarction in the appropriate clinical setting. Current methods to diagnose PE are not only expensive but lack specificity and sensitivity, and the misdiagnosis of this condition is a leading cause of preventable death in hospitalized patients (Raskob G.E. and Hull R.D., 1999, Curr Opin Hematol. 6(5): 280-4).

In another embodiment, hCLASP-2-based diagnostics involve screening assays for identifying disorders of cells of hematopoietic lineage. hCLASP-2 is expressed in human T cells, B cells but not cells from the myeloid lineage. Different hCLASP-2 isoforms in T and B cells permit further discrimination between malignancies of T and B lineage (FIG. 4B). Precise identification of hematopoietic cell types is vital to guide chemotherapy and radiation therapy of patients with leukemia and lymphoma (Fauci et al Eds., 1998, Harrison's Principles of Internal Medicine, 14th Ed. McGraw Hill, pp. 695-712). hCLASP-2 expression differences can be detected by using FACS, immunofluorescence, immunoperoxidase staining, RT-PCR, in situ hybridization or RNA blot analysis (Sambrook, Fritsch and Maniatas, Molecular Cloning, 2nd Ed. Cold Spring Harbor Lab. Press, 1989; Ward MS, Pathology 1999 Nov; 31(4): 382-92).

In another embodiment, hCLASP-2-based diagnostics involve screening assays for identifying activated immune system cells. Although hCLASP-2 is generally expressed at quite low levels in PBMCs (which is critical for some of the above applications), it is known that the surface expression of the closely related mouse CLASP-1 protein is altered during the process of lymphocyte activation. An analogous change in expression is expected for the hCLASP-2 protein. Subtyping lymphocytes specific for a particular antigen, for example, using MHC-based multimeric staining reagents (Altman et. al., 1996, Science 274: 94-6), for separating cell populations into hCLASP-2 high and hCLASP-2 low populations, can aid in determining the nature of the immune response against that antigen.

Such understanding is critical, for example, in predicting the course of chronic viral infections such as hepatitis B, hepatitis C, and HIV, and to designing appropriate treatment regimens for patients suffering from these infections.

hCLASP-2 can also serve as a potential therapeutic agent for Wilms' tumor. Wilms tumor is the most common primary renal tumor of childhood (Cotran, Kumar, and Collins, 1999, Robbins Pathologic Basis of Disease, 6th Ed. W.B. Saunders, pp. 487-89). As discussed herein, hCLASP-2 is highly expressed in 293 cells, embryonic kidney epithelial cells. Therefore, hCLASP-2 nucleic or amino acid sequence or fragments can serve as tumor markers for Wilms' tumor. Antibodies directed against a hCLASP-2 variant that is expressed only in Wilms' tumor can serve as novel therapeutic agents for Wilms' tumor, and can also function as delivery vehicles for other targeted therapeutics that may be attached to the anti-hCLASP-2 antibody (*e.g.*, chemotherapeutics or radiolabeling).

5.4.2.2.1. CLASP-2 Antisense, Ribozyme and Triplex Polynucleotides and Methods of Use

Oligonucleotide sequences, that include anti-sense RNA and DNA molecules and ribozymes that function to inhibit the translation of a CLASP-2 mRNA are within the scope of the invention. Such molecules are useful in cases where downregulation of CLASP-2 expression is desired. Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by binding to targeted mRNA and preventing protein translation. The invention provides methods and antisense oligonucleotide or polynucleotide reagents which can be used to reduce expression of CLASP-2 gene products in vitro or in vivo. Administration of the antisense reagents of the invention to a target cell results in reduced CLASP activity. As will be apparent to one of skill and as discussed *supra* (Table 3), specific CLASP-2 splice variants can be specifically targeted for inhibition. Alternatively, by designing an, *e.g.*, antisense molecule that recognizes a sequence found in several or all CLASP-2 species, a general inhibition can be achieved.

A. Antisense

Without intending to be limited to any particular mechanism, it is believed that antisense oligonucleotides bind to, and interfere with the translation of, the sense CLASP-2 mRNA. Alternatively, the antisense molecule can render the CLASP-2 mRNA susceptible to nuclease digestion, interfere with transcription, interfere with processing, localization or otherwise with RNA precursors ("pre-mRNA"), repress transcription of mRNA from the

CLASP-2 gene, or act through some other mechanism. However, the particular mechanism by which the antisense molecule reduces CLASP-2 expression is not critical.

The antisense polynucleotides of the invention comprise an antisense sequence of at least 7 to 10 to typically 20 or more nucleotides that specifically hybridize to a sequence from mRNA encoding CLASP-2 or mRNA transcribed from the CLASP-2 gene. More often, the antisense polynucleotide of the invention is from about 10 to about 50 nucleotides in length or from about 14 to about 35 nucleotides in length. In other embodiments, antisense polynucleotides are polynucleotides of less than about 100 nucleotides or less than about 200 nucleotides. In general, the antisense polynucleotide should be long enough to form a stable duplex but short enough, depending on the mode of delivery, to administer in vivo, if desired. The minimum length of a polynucleotide required for specific hybridization to a target sequence depends on several factors, such as G/C content, positioning of mismatched bases (if any), degree of uniqueness of the sequence as compared to the population of target polynucleotides, and chemical nature of the polynucleotide (*e.g.*, methylphosphonate backbone, peptide nucleic acid, phosphorothioate), among other factors. Generally, to assure specific hybridization, the antisense sequence is substantially complementary to the target CLASP-2 mRNA sequence. In certain embodiments, the antisense sequence is exactly complementary to the target sequence. The antisense polynucleotides can also include, however, nucleotide substitutions, additions, deletions, transitions, transpositions, or modifications, or other nucleic acid sequences or non-nucleic acid moieties so long as specific binding to the relevant target sequence corresponding to CLASP-2 RNA or its gene is retained as a functional property of the polynucleotide.

It will be appreciated that the CLASP-2 polynucleotides and oligonucleotides of the invention can be made using nonstandard bases (*e.g.*, other than adenine, cytidine, guanine, thymine, and uridine) or nonstandard backbone structures to provides desirable properties (*e.g.*, increased nuclease-resistance, tighter-binding, stability or a desired TM). Techniques for rendering oligonucleotides nuclease-resistant include those described in PCT publication WO 94/12633. A wide variety of useful modified oligonucleotides may be produced, including oligonucleotides having a peptide-nucleic acid (PNA) backbone (Nielsen *et al.*, 1991, Science 254: 1497) or incorporating 2'-O-methyl ribonucleotides, phosphorothioate nucleotides, methyl phosphonate nucleotides, phosphotriester nucleotides, phosphorothioate nucleotides, phosphoramidates. Still other useful oligonucleotides may contain alkyl and halogen-substituted sugar moieties comprising one of the following at the 2' position: OH, SH, SCH3, F, OCN, OCH3OCH3, OCH3O(CH2)nCH3, O(CH2)nNH2 or

O(CH2)nCH3, where n is from 1 to about 10; C1 to C10 lower alkyl, substituted lower alkyl, alkaryl or aralkyl; Cl; Br; CN; CF3; OCF3; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; SOCH3 ; SO2CH3; ONO2; NO2; N3; NH2; heterocycloalkyl; heterocycloalkaryl; aminoalkylamino; polyalkylamino; substituted silyl; an RNA cleaving group; a cholesteryl group; a folate group; a reporter group; an intercalator; a group for improving the pharmacokinetic properties of an oligonucleotide; or a group for improving the pharmacodynamic properties of an oligonucleotide and other substituents having similar properties. Folate, cholesterol or other groups that facilitate oligonucleotide uptake, such as lipid analogs, may be conjugated directly or via a linker at the 2' position of any nucleoside or at the 3' or 5' position of the 3'-terminal or 5'-terminal nucleoside, respectively. One or more such conjugates may be used.

Oligonucleotides may also have sugar mimetics such as cyclobutyls in place of the pentofuranosyl group. Other embodiments may include at least one modified base form or "universal base" such as inosine, or inclusion of other nonstandard bases such as queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases. The antisense oligonucleotide can comprise at least one modified base moiety which is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

The invention further provides oligonucleotides having backbone analogues such as phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, chiral-methyl phosphonates, nucleotides with short chain alkyl or cycloalkyl intersugar linkages, short chain heteroatomic or heterocyclic

intersugar ("backbone") linkages, or CH2-NH-O-CH2, CH2-N(CH3)-OCH2, CH2-O-N(CH3)-CH2, CH2-N(CH3)-N(CH3)-CH2 and O-N(CH3)-CH2-CH2 backbones (where phosphodiester is O-P-O-CH2), or mixtures of the same. Also useful are oligonucleotides having morpholino backbone structures (U.S. Patent No. 5,034,506).

Useful references include Oligonucleotides and Analogues, A Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan *et al.*, 9 July 1993, J. Med. Chem. 36(14): 1923-1937; Antisense Research and Applications (1993, CRC Press), in its entirety and specifically Chapter 15, by Sanghvi, entitled "Heterocyclic base modifications in nucleic acids and their applications in antisense oligonucleotides;" and Antisense Therapeutics, ed. Sudhir Agrawal (Humana Press, Totowa, New Jersey, 1996).

In one embodiment, the antisense sequence is complementary to relatively accessible sequences of the CLASP-2 mRNA (*e.g.*, relatively devoid of secondary structure). This can be determined by analyzing predicted RNA secondary structures using, for example, the MFOLD program (Genetics Computer Group, Madison WI) and testing in vitro or in vivo as is known in the art. Another useful method for identifying effective antisense compositions uses combinatorial arrays of oligonucleotides (see, *e.g.*, Milner *et al.*, 1997, Nature Biotechnology 15: 537). Examples of oligonucleotides that can be tested in cells for antisense suppression of CLASP-2 function are those capable of hybridizing to (*i.e.*, substantially complementary to) the following positions from SEQUENCE ID NO: 1:

1) GAAGGCGATCATCACGTGGCCTTCCATCGC
2) GCTTCAAGTAATGACTGGTGCAGAACATCTG
3) GCTCCTCCTCAGGCAGGCGCTATGGCTGTGG
4) GTAGGCCCGGTGCAGCGTGTCATACAGATGG

(See also Example 8)

In some embodiments, administration of antisense oligonucleotides will result in reduction of hCLASP-mRNA expression by at least about 50%, as assessed by Northern analysis after administration of an antisense phosphorothioate oligonucleotide at a concentration of 1 μM, 5 μM, 10 μM or 20 μM.

The invention also provides an antisense polynucleotide that has sequences in addition to the antisense sequence (*i.e.*, in addition to anti-CLASP-2-sense sequence). In this

case, the antisense sequence is contained within a polynucleotide of longer sequence. In another embodiment, the sequence of the polynucleotide consists essentially of, or is, the antisense sequence.

The antisense nucleic acids (DNA, RNA, modified, analogues, and the like) can be made using any suitable method for producing a nucleic acid, such as the chemical synthesis and recombinant methods disclosed herein. In one embodiment, for example, antisense RNA molecules of the invention can be prepared by *de novo* chemical synthesis or by cloning. For example, an antisense RNA that hybridizes to CLASP-2 mRNA can be made by inserting (ligating) an CLASP-2 DNA sequence (*e.g.*, SEQUENCE ID No: 1, or fragment thereof) in reverse orientation operably linked to a promoter in a vector (*e.g.*, plasmid). Provided that the promoter and, preferably termination and polyadenylation signals, are properly positioned, the strand of the inserted sequence corresponding to the noncoding strand will be transcribed and act as an antisense oligonucleotide of the invention. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter or enhancer) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

In one embodiment, antisense DNA oligodeoxyribonucleotides derived from the translation initiation site, *e.g.*, between -10 and +10 regions of a CLASP-2 nucleotide sequence, are used. For general methods relating to antisense polynucleotides, see ANTISENSE RNA AND DNA, 1988, D.A. Melton, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). See also, Dagle *et al.*, 1991, Nucleic Acids Research, 19: 1805. For a review of antisense therapy, see, *e.g.*, Uhlmann *et al.*, 1990, Chem. Reviews, 90: 543-584.

B. Ribozyme

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of CLASP-2 RNA sequences.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing

the cleavage site can be evaluated for predicted structural features such as secondary structure that can render the oligonucleotide sequence unsuitable. The suitability of candidate targets can also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

C. **Triplex**

Alternatively, endogenous target gene expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of the target gene (*i.e.*, the target gene promoter and/or enhancers) to form triple helical structures that prevent transcription of the target gene in target cells in the body. (See generally, Helene, 1991, Anticancer Drug Des., 6(6): 569-584; Helene *et al.*, 1992, Ann. N.Y. Acad. Sci., 660: 27-36; and Maher, 1992, Bioassays 14(12): 807-815).

Nucleic acid molecules to be used in triplex helix formation for the inhibition of transcription should be single stranded and composed of deoxynucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences can be pyrimidine-based, which will result in TAT and CGC^+ triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarily to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules can be chosen that are purine-rich, for example, contain a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation can be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizable stretch of either purines or pyrimidines to be present on one strand of a duplex.

D. **General**

The anti-sense RNA and DNA molecules, ribozymes and triple helix molecules of the invention can be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as for example solid phase

phosphoramidite chemical synthesis. Alternatively, RNA molecules can be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences can be incorporated into a wide variety of vectors which contain suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various modifications to the DNA molecules can be introduced as a means of increasing intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences of ribo- or deoxy- nucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Methods for introducing polynucleotides into such cells or tissue include methods for *in vitro* introduction of polynucleotides such as the insertion of naked polynucleotide, *i.e.*, by injection into tissue, the introduction of a CLASP-2 polynucleotide in a cell *ex vivo*, the use of a vector such as a virus, (*e.g.*, a retrovirus, adenovirus, adeno-associated virus, and the like), phage or plasmid, and the like or techniques such as electroporation or calcium phosphate precipitation.

5.4.2.2.2. Gene Therapy

By introducing gene sequences into cells, gene therapy can be used to treat conditions in which the cells do not express normal CLASP-2 or express abnormal/inactive CLASP-2. In some instances, the polynucleotide encoding a CLASP-2 is intended to replace or act in the place of a functionally deficient endogenous gene. Alternatively, abnormal conditions characterized by overexpression can be treated using the gene therapy techniques described below.

In a specific embodiment, nucleic acids comprising a sequence encoding a CLASP-2 protein or functional derivative thereof, are administered to promote CLASP-2 function, by way of gene therapy. Gene therapy refers to therapy performed by the administration of a nucleic acid to a subject. In this embodiment of the invention, the nucleic acid produces its encoded protein that mediates a therapeutic effect by promoting CLASP-2 function.

Any of the methods for gene therapy available in the art can be used according to the present invention. Exemplary methods are described below.

For general reviews of the methods of gene therapy, *see*, Goldspiel *et al.*, 1993, Clinical Pharmacy 12: 488-505; Wu and Wu, 1991, Biotherapy 3: 87-95; Tolstoshev, 1993, Ann. Rev. Pharmacol. Toxicol. 32: 573-596; Mulligan, 1993, Science 260: 926-932; and Morgan and Anderson, 1993, Ann. Rev. Biochem. 62: 191-217; Can, 1993, TIBTECH 11(5): 155-215). Methods commonly known in the art of recombinant DNA technology which can be used are described in Ausubel *et al.*, *supra*; and Kriegler, 1990, Gene Transfer and Expression, A Laboratory Manual, Stockton Press, NY.

In one aspect, the therapeutic composition comprises a CLASP-2 nucleic acid that is part of an expression vector that encodes a CLASP-2 protein or fragment or chimeric protein thereof in a suitable host. In particular, such a nucleic acid has a promoter operably linked to the CLASP-2 coding region, said promoter being inducible or constitutive, and, optionally, tissue-specific. In another particular embodiment, a nucleic acid molecule is used in which the CLASP-2 coding sequences and any other desired sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of the CLASP-2 nucleic acid (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. U.S.A. 86: 8932-8935; Zijlstra *et al.*, 1989, *Nature* 342: 435-438).

Delivery of the nucleic acid into a patient can be either direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying vector, or indirect, in which case, cells are first transformed with the nucleic acid *in vitro*, then transplanted into the patient. These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy.

In a specific embodiment, the nucleic acid is directly administered *in vivo*, where it is expressed to produce the encoded product. This can be accomplished by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by infection using a defective or attenuated retroviral or other viral vector (*see*, U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, or by administering it in linkage to a peptide which is known to enter the nucleus, by administering it in linkage to a ligand subject to receptor-mediated endocytosis (*see, e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262: 4429-4432) (which can be used to target cell types specifically expressing the receptors), and the like. In another embodiment, a nucleic acid-ligand

complex can be formed in which the ligand comprises a fusogenic viral peptide to disrupt endosomes, allowing the nucleic acid to avoid lysosomal degradation. In yet another embodiment, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (*see, e.g.*, PCT Publications WO 92/06180 dated April 16, 1992; WO 92/22635 dated December 23, 1992; WO 92/20316 dated November 26, 1992; WO 93/14188 dated July 22, 1993; WO 93/20221 dated October 14, 1993). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. U.S.A. 86: 8932-8935; Zijlstra *et al.*, 1989, Nature 342: 435-438).

In a specific embodiment, a viral vector that contains the CLASP-2 nucleic acid is used. For example, a retroviral vector can be used (*see*, Miller *et al.*, 1993, Meth. Enzymol. 217: 581-599). These retroviral vectors have been modified to delete retroviral sequences that are not necessary for packaging of the viral genome and integration into host cell DNA. The CLASP-2 nucleic acid to be used in gene therapy is cloned into the vector, which facilitates delivery of the gene into a patient. More detail about retroviral vectors can be found in Boesen *et al.*, 1994, Biotherapy 6: 291-302, which describes the use of a retroviral vector to deliver the *mdr1* gene to hematopoietic stem cells in order to make the stem cells more resistant to chemotherapy. Other references illustrating the use of retroviral vectors in gene therapy are: Clowes *et al.*, 1994, J. Clin. Invest. 93: 644-651; Kiem *et al.*, 1994, Blood 83: 1467-1473; Salmons and Gunzberg, 1993, Human Gene Therapy 4: 129-141; and Grossman and Wilson, 1993, Curr. Opin. in Genetics and Devel. 3: 110-114.

Adenoviruses are other viral vectors that can be used in gene therapy. Adenoviruses are especially attractive vehicles for delivering genes to respiratory epithelia. Adenoviruses naturally infect respiratory epithelia where they cause a mild disease. Other targets for adenovirus-based delivery systems are liver, the central nervous system, endothelial cells, and muscle. Adenoviruses have the advantage of being capable of infecting non-dividing cells. Kozarsky and Wilson 1993, Current Opinion in Genetics and Development 3: 499-503) present a review of adenovirus-based gene therapy. Bout *et al.*, 1994, Human Gene Therapy 5: 3-10, demonstrated the use of adenovirus vectors to transfer genes to the respiratory epithelia of rhesus monkeys. Other instances of the use of adenoviruses in gene therapy can be found in Rosenfeld *et al.*, 1991, Science 252: 431-434; Rosenfeld *et al.*, 1992, Cell 68: 143-155; and Mastrangeli *et al.*, 1993, J. Clin. Invest. 91: 225-234. Adeno-associated virus (AAV) has also been proposed for use in gene therapy (Walsh *et al.*, 1993, Proc. Soc. Exp. Biol. Med. 204: 289-300.

Another approach to gene therapy involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the nucleic acid is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the nucleic acid sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, and the like. Numerous techniques are known in the art for the introduction of foreign genes into cells (*see, e.g.*, Loeffler and Behr, 1993, Meth. Enzymol. 217: 599-618; Cohen *et al.*, 1993, Meth. Enzymol. 217: 618-644; Cline, 1985, Pharmac. Ther. 29: 69-92) and can be used in accordance with the present invention, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the nucleic acid to the cell, so that the nucleic acid is expressible by the cell and preferably heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. In a preferred embodiment, epithelial cells are injected, *e.g.*, subcutaneously. In another embodiment, recombinant skin cells can be applied as a skin graft onto the patient. Recombinant blood cells (*e.g.*, hematopoietic stem or progenitor cells) are preferably administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, and the like., and can be determined by one skilled in the art.

Cells into which a nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, *e.g.*, as obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, and the like. In a preferred embodiment, the cell used for gene therapy is autologous to the patient.

In a specific embodiment, the nucleic acid to be introduced for purposes of gene therapy comprises an inducible promoter operably linked to the coding region, such that expression of the nucleic acid is controllable by controlling the presence or absence of the appropriate inducer of transcription.

5.4.2.3. Knockout Cells

In one aspect of the invention, endogenous target gene expression can also be reduced by inactivating or "knocking out" the target gene or its promoter using targeted homologous recombination (*see, e.g.*, Smithies *et al.*, 1985, Nature 317: 230-234; Thomas and Capecchi, 1987, Cell 51: 503-512; Thompson *et al.*, 1989, Cell 5: 313-321; each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional target gene (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous target gene (either the coding regions or regulatory regions of the target gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express the target gene *in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the target gene. Such approaches are particularly suited in the agricultural field where modifications to ES (embryonic stem) cells can be used to generate animal offspring with an inactive target gene (*see, e.g.*, Thomas and Capecchi, 1987 and Thompson, 1989, *supra*). However, this approach can be adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors.

5.4.2.4. Transgenic and Knockout Animals

The CLASP-2 gene product can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, sheep, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees can be used to generate CLASP-2 transgenic animals. The term "transgenic," as used herein, refers to animals expressing CLASP-2 gene sequences from a different species (*e.g.*, mice expressing human CLASP-2 gene sequences), as well as animals that have been genetically engineered to overexpress endogenous (*i.e.*, same species) CLASP-2 sequences or animals that have been genetically engineered to no longer express endogenous CLASP-2 gene sequences (*i.e.*, "knock-out" animals), and their progeny.

Any technique known in the art can be used to introduce a CLASP-2 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but

are not limited to pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci., U.S.A. 82: 6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56: 313-321); electroporation of embryos (Lo, 1983, Mol. Cell. Biol. 3: 1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57: 717-723) (For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. 115, 171-229)

Any technique known in the art can be used to produce transgenic animal clones containing a CLASP-2 transgene, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal or adult cells induced to quiescence (Campbell *et al.*, 1996, Nature 380: 64-66; Wilmut *et al.*, Nature 385: 810-813).

The present invention provides for transgenic animals that carry a CLASP-2 transgene in all their cells, as well as animals that carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene can be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene can also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko *et al.* (1992, Proc. Natl. Acad. Sci. U.S.A. 89: 6232-6236). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the CLASP-2 transgene be integrated into the chromosomal site of the endogenous CLASP-2 gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous CLASP-2 gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous CLASP-2 gene. The transgene can also be selectively introduced into a particular cell type, thus inactivating the endogenous CLASP-2 gene in only that cell type, by following, for example, the teaching of Gu *et al.* (1994, Science 265: 103-106). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant CLASP-2 gene can be assayed utilizing standard techniques. Initial screening can be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals can also be assessed using techniques

that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR (reverse transcriptase PCR). Samples of CLASP-2 gene-expressing tissue, can also be evaluated immunocytochemically using antibodies specific for the CLASP-2 transgene product.

5.4.2.5. Other Uses of CLASP-2 Polynucleotides

There exists an ongoing need to identify new chromosome marking reagents. Sequences can be mapped to chromosomes by preparing PCR primers from SEQ ID NO: 1, 3, 5, or 9. These primers can be can be less than 50 nucleotides in length, generally less than 46 nucleotides, more generally less than 41 nucleotides, most generally less than 36 nucleotides, preferably less than 31 nucleotides, more preferably less than 26 nucleotides, and most preferably less than 21 nucleotides in length. The probes can also be less than 16 nucleotides, less than 13 nucleotides in length, less than 9 nucleotides in length and less than 7 nucleotides in length. Primers can be selected so that the primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes (*i.e.*, chromosome 13). Only those hybrids containing the human CLASP-2 gene corresponding to SEQ ID NO: 1, 3, 5, or 9 will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Precise chromosomal location of the CLASP-2 polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. See Verma, et al, Human Chromosomes: A Manual of Basic Techniques, Pergamon Press. NY, 1988. Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. See McKusick, V., 1998, Mendelian Inheritance in Man : A Catalog of Human Genes and Genetic Disorders, 12th Ed, Johns Hopkins University Press.

The CLASP-2 polynucleotides can be used for identifying individuals from minute biological samples as DNA markers for restriction fragment length polymorphism (RFLP). An individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot with CLASP-2 DNA markers to yield unique bands for identifying the individual.

As described above, upon sequencing of numerous independent cDNA products, single nucleotide polymorphisms (SNPs) have been discovered within CLASP-2. These alterations and differences are presented in FIG 11B. They represent mis-sense alterations.

If it is determined that certain SNPs are deleterious or advantageous, SNPs can be used as a diagnostic tool through SNP mapping or direct sequencing of the SNP region to determine which isoform is expressed. Additionally, the SNPs can be used as a general SNP marker for chromosomal defects such as rearrangement and translocations.

CLASP-2 polynucleotides can be also be used as polymorphic markers for forensic analysis. See generally National Research Council, The Evaluation of Forensic DNA Evidence (Eds. 1996, Pollard *et al.*, National Academy Press, Washington D.C.). The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (*e.g.*, by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample. The CLASP-2 polynucleotide sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO: 1, 3, 5 or 9 are particularly appropriate for this use as greater numbers of polymorphisms occur in the

noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the CLASP-2 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO: 1, 3, 5, or 9 having a length of at least 20 bases, preferably at least 25 bases, and more preferably at least 30 bases.

CLASP-2 polynucleotides can also be used as reagents for paternity testing. The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child. Of course, the present invention can be expanded to the use of this procedure to determine if one individual is related to another. Even more broadly, the present invention can be employed to determine how related one individual is to another, for example, between races or species.

Bacterial infections are a major cause of health-related problems. However, the emergence of drug resistant bacteria is compromising the therapeutic value of the present spectrum of antibiotics. All the currently used antibiotics are small organic molecules, with certain level of structural similarity. This provides an advantage for bacteria to develop drug resistance, since they need to modify a limited number of genes in order to become resistant to a wide variety of antibiotics. The development of antibiotics with different chemical structure and targets can overcome antibiotic resistance, and provide therapeutic superiority in preventing infection by bacterial pathogens. Additionally, most antibiotics are not naturally occurring compounds and cause minor or sometimes serious side effects. For example, antibiotics used to treat TB can cause hearing loss.

The present invention provides new antibacterial agents. Certain CLASP-2 DNA sequences were difficult to clone and subclone (*see* Example 1). Bacteria harboring certain pieces of CLASP cDNA products were unable to be isolated, indicating that introduction of CLASP sequencescompromised bacterial viability. There can be at least two possible reasons why the CLASP cDNA were unable to be cloned, which can reflect a variation of the well-established Modification and Restriction systems found in bacteria (reviewed in Wilson and Murray. (1991) Annu. Rev. Genet. 25:585-627; Bickle and Kruger (1993) Microbiol. Rev. 57:29-67). This well-described system is used by bacteria to prevent deleterious effects caused by the introduction of foreign DNA. Bacteria can recognize

foreign DNA since it does not have the same modifications (e.g. methylation) as the native DNA. After recognition, the bacteria then digest and eliminate the foreign DNA (restriction). In the first scenario, the CLASP cDNA can be recognized as foreign DNA, and digested and eliminated as in the Modification and Restriction system. However, this would be unique for CLASP cDNA since the bacteria used for cloning cDNA are compromised in the Modification and Restriction system, which makes cloning of cDNA into bacteria a practice common in the art. If this is the case, the bacterial apparatus that specifically recognizes or eliminates CLASP cDNA can provide a novel target to develop antimicrobial agents. The CLASP DNA sequence would be useful in targeting the apparatus as well as an entry point for designing screens to identify potential targets. The second possibility is that CLASP cDNA behaves as an antimicrobial agent (i.e., antibiotic), and prevents bacterial growth. This, in effect, would create a new type of antibiotic mediated by the presence of foreign DNA (i.e. CLASP cDNA). In the case for the CLASP cDNA, the bacteria can recognize the DNA but instead of digesting and eliminating the DNA, the CLASP cDNA can cause a variation of the restriction and prevent the bacteria from growing, imposing a bacteriacidal effect upon the bacteria.

DNA as an antimicrobial agent has significant advantages over currently available agents. First, it is structurally unrelated to any existing antibiotics, and can overcome the present growing drug-resistance problem to structurally common agents. Second, since DNA antimicrobials composed of naturally-occurring human DNA, are expected to have minimal side effects and immune rejection. Third, DNA sequences can be tailored with sequence variation and numerous chemical modifications to circumvent the problem of resistance. Fourth, the antimicrobial DNA can be delivered specifically to bacterial cells through the use of bacteriophages (i.e., bacterial virus) which specifically infect bacteria and do not infect human cells. Further specificity can be generated to infect certain bacteria and bacterial subpopulations. Finally, this system can be economically robust since the generation of DNA and delivery vehicles are inexpensive.

5.5. Polypeptides Encoded by the CLASP-2 Gene Coding Sequence

In accordance with the invention, a CLASP-2 polynucleotide which encodes the CLASP-2 polypeptides, mutant polypeptides, peptide fragments, CLASP-2 fusion proteins or functional equivalents thereof, can be used to express CLASP-2 proteins in appropriate host cells. In various embodiments, the CLASP-2 polypeptides expressed will be identical or substantially similar to SEQ ID NOs: 2, 4, 6 or 10 or a fragment thereof.

In some embodiments, altered DNA sequences which can be used in accordance with the invention include deletions, additions or substitutions of different nucleotide residues resulting in a sequence that encodes the same or a functionally equivalent gene product. For example, due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence, can be used in the practice of the invention for the expression of the CLASP-2 protein. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. One of skill will recognize that each codon in a nucleic acid sequence such SEQ ID NO: 1 (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence. Thus, for example, due to the degeneracy of the genetic code, a polypeptide having the sequence of SEQ ID NO: 2 or a fragment thereof, can be encoded by numerous polynucleotides other than SEQ ID NO: 1. Typically, the degenerate sequence will hybridize with SEQ ID NO: 1 under high or moderate stringency conditions, but this is not strictly required (*e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions.)

The gene product itself can contain deletions, additions or substitutions of amino acid residues within a CLASP-2 sequence, which result in a silent change thus producing a functionally equivalent CLASP-2 protein. Such conservative amino acid substitutions can be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine, histidine and arginine; amino acids with uncharged polar head groups having similar hydrophilicity values include the following: glycine, asparagine, glutamine, serine, threonine, tyrosine; and amino acids with nonpolar head groups include alanine, valine, isoleucine, leucine, phenylalanine, proline, methionine, tryptophan. Creighton, 1984, PROTEINS, has grouped amino acids that are conservative substitutions for

one another as follows: (1) Alanine (A), Glycine (G); (2) Aspartic acid (D), Glutamic acid (E); (3) Asparagine (N), Glutamine (Q); (4) Arginine (R), Lysine (K); (5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); (6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (7) Serine (S), Threonine (T); and (8) Cysteine (C), Methionine (M).

The DNA sequences of the invention can be engineered in order to alter a CLASP-2 coding sequence for a variety of ends, including but not limited to, alterations which modify processing and expression of the gene product. For example, mutations can be introduced using techniques which are well known in the art, *e.g.*, site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, phosphorylation, and the like. Based on the domain organization of the CLASP-2 proteins, a large number of CLASP-2 mutant polypeptides can be constructed by modifying or rearranging the nucleotide sequences that encode the CLASP-2 extracellular, transmembrane and cytoplasmic domains.

In various embodiments, the present invention provides homologues of the CLASP-2 polypeptides which function as either an CLASP-2 agonists or an CLASP-2 antagonist. In a preferred embodiment, the CLASP-2 agonists and antagonists stimulate or inhibit, respectively, a subset of the biological activities of the naturally occurring form of the CLASP-2 polypeptide. Thus, specific biological effects can be elicited by treatment with a homologue of limited function. In one embodiment, treatment of a subject with a homologue having a subset of the biological activities of the naturally occurring form of the polypeptide has fewer side effects in a subject relative to treatment with the naturally occurring form of the CLASP-2 polypeptide.

The invention contemplates both full-length CLASP-2 polypeptides and fragments, *e.g.*, fragments having a length of at least about 10, often 20, frequently 50 or 100 residues substantially identical to the exemplified CLASP-2 polypeptide sequences of the invention. Protein fragments can be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 2 1-40, 4 1-60, 61-80, 81-100, 102-120, 121-140, 141-160, 161-180, 181-200, or 201 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the CLASP-2 protein. Further preferred polypeptide fragments include the CLASP-2 protein having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-X, can be deleted from the amino terminus of either the CLASP-2 polypeptide. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these CLASP-2 polypeptide fragments are also preferred.

Even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities can still be retained. Thus, the ability of shortened CLASP-2 muteins to induce and/or bind to antibodies which recognize the complete or mature forms of the polypeptides generally will be retained when less than the majority of the residues of the complete or mature polypeptide are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete polypeptide retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a CLASP-2 mutein with a large number of deleted N-terminal amino acid residues can retain some biological or immunogenic activities. In fact, peptides composed of as few as four CLASP-2 amino acid residues can often evoke an immune response.

Homologues of the CLASP-2 polypeptide can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the CLASP-2 polypeptide. As used herein, the term "homologue" refers to a variant form of the CLASP-2 polypeptide which acts as an agonist or antagonist of the activity of the CLASP-2 polypeptide. An agonist of the CLASP-2 polypeptide can retain substantially the same, or a subset, of the biological activities of the CLASP-2 polypeptide. An antagonist of the CLASP-2 polypeptide can inhibit one or more of the activities of the naturally occurring form of the CLASP-2 polypeptide, by, for example, competitively binding to a downstream or upstream member of the CLASP-2 molecular pathway which includes the CLASP-2 polypeptide.

Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, *e.g.*, changes in RNA or protein levels, changes in protein activity, changes in product levels, changes in downstream gene expression, changes in reporter gene transcription (luciferase, CAT, β-galactosidase, β-glucuronidase, GFP (*see, e.g.*, Mistili & Spector, 1997, Nature Biotechnology 15: 961-964); changes in signal transduction, phosphorylation and

dephosphorylation, receptor-ligand interactions, second messenger concentrations (*e.g.*, cGMP, cAMP, IP_3, and Ca^{2+}), and cell growth. These assays can be *in vitro*, *in vivo*, and *ex vivo*. Such functional effects can be measured by any means known to those skilled in the art, *e.g.*, measurement of RNA or protein levels, measurement of RNA stability, identification of downstream or reporter gene expression, *e.g.*, via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP_3); changes in intracellular calcium levels; cytokine release, and the like.

5.5.1. Synthesis or Expression of CLASP-2 Polypeptide Expression Systems

In order to express a biologically active CLASP-2, the nucleotide sequence coding for CLASP-2, or a functional equivalent, is inserted into an appropriate expression vector. The CLASP-2 gene product as well as host cells or cell lines transfected or transformed with recombinant CLASP-2 expression vectors can be used for a variety of purposes. These include, but are not limited to, generating antibodies (*i.e.*, monoclonal or polyclonal) that competitively inhibit activity of CLASP-2 protein and neutralize its activity; antibodies that activate CLASP-2 function and antibodies that detect its presence on the cell surface or in solution. Anti-CLASP-2 antibodies can be used in detecting and quantifying expression of CLASP-2 levels in cells and tissues such as lymphocytes and macrophages, as well as isolating CLASP-2-positive cells from a cell mixture.

Methods which are well known to those skilled in the art can be used to construct recombinant expression vectors containing the CLASP-2 coding sequence and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. (*See, e.g.*, the techniques described in Sambrook *et al.*, 1989, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. and Ausubel *et al.*, *supra*). The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides or

peptides, including fusion polypeptides or peptides, encoded by nucleic acids as described herein (*e.g.*, CLASP-2 polypeptides, mutant forms of CLASP-2, fusion polypeptides, and the like).

A variety of host-expression vector systems can be utilized to express a CLASP-2 coding sequence. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the CLASP-2 coding sequence; yeast transformed with recombinant yeast expression vectors containing the CLASP-2 coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the CLASP-2 coding sequence; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing the CLASP-2 coding sequence; or animal cell systems. The expression elements of these systems vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of suitable transcription and translation elements, including constitutive and inducible promoters, can be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage λ, plac, ptrp, ptac (ptrp-lac hybrid promoter; cytomegalovirus promoter) and the like can be used; when cloning in insect cell systems, promoters such as the baculovirus polyhedron promoter can be used; when cloning in plant cell systems, promoters derived from the genome of plant cells (*e.g.*, heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll α/β binding protein) or from plant viruses (*e.g.*, the ^{35}S RNA promoter of CaMV; the coat protein promoter of TMV) can be used; when cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter) can be used; when generating cell lines that contain multiple copies of the CLASP-2 DNA, SV40-, BPV- and EBV-based vectors can be used with an appropriate selectable marker.

In bacterial systems a number of expression vectors can be advantageously selected depending upon the use intended for the expressed CLASP-2 product. For example, when large quantities of CLASP-2 protein are to be produced for the generation of antibodies or to screen peptide libraries, vectors which direct the expression of high levels of fusion protein products that are readily purified can be desirable. Such vectors include, but are not

limited to, the E. coli expression vector pUR278 (Ruther *et al.*, 1983, EMBO J. 2: 1791), in which the CLASP-2 coding sequence can be ligated into the vector in frame with the lacZ coding region so that a hybrid protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic acids Res. 13: 3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264: 5503-5509); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In yeast, a number of vectors containing constitutive or inducible promoters can be used. (Current Protocols in Molecular Biology, Vol. 2, 1988 (Suppl. 1999), Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant *et al.*, 1987, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 1987, Acad. Press, N.Y., Vol. 153, pp. 516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; and The Molecular Biology of the Yeast Saccharomyces, 1982, Eds. Strathern *et al.*, Cold Spring Harbor Press, Vols. I and II.)

In cases where plant expression vectors are used, the expression of the CLASP-2 coding sequence can be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA promoters of CaMV (Brisson *et al.*, 1984, Nature 310: 511-514), or the coat protein promoter of TMV (Takamatsu *et al.*, 1987, EMBO J. 6: 307-311) can be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.*, 1984, EMBO J. 3: 1671-1680; Broglie *et al.*, 1984, Science 224: 838-843); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley *et al.*, 1986, Mol. Cell. Biol. 6: 559-565) can be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, and the like. (Weissbach & Weissbach, 1988, Methods for Plant Molecular Biology, Academic Press, NY, Section VIII, pp. 421-463; and Grierson & Corey, 1988, Plant Molecular Biology, 2d Ed., Blackie, London, Ch. 7-9.)

An alternative expression system which could be used to express CLASP-2 is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera*

frugiperda cells. The CLASP-2 coding sequence can be cloned into non-essential regions (*e.g.*, the polyhedron gene) of the virus and placed under control of an AcNPV promoter (*e.g.*, the polyhedron promoter). Successful insertion of the *CLASP-2* coding sequence will result in inactivation of the polyhedron gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedron gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*see, e.g.*, Smith *et al.*, 1983, J. Viol. 46: 584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral based expression systems can be utilized. In cases where an adenovirus is used as an expression vector, the CLASP-2 coding sequence can be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene can then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing CLASP-2 in infected hosts. (*See, e.g.*, Logan & Shenk, 1984, Proc. Natl. Acad. Sci. U.S.A. 81: 3655-3659). Alternatively, the vaccinia 7.5K promoter can be used. (*See, e.g.*, Mackett *et al.*, 1982, Proc. Natl. Acad. Sci. U.S.A. 79: 7415-7419; Mackett *et al.*, 1984, J. Virol. 49: 857-864; Panicali *et al.*, 1982, Proc. Natl. Acad. Sci. U.S.A. 79: 4927-4931). Regulatable expression vectors such as the tetracycline repressible vectors can also be used to express a coding sequence in a controlled fashion.

Specific initiation signals can also be required for efficient translation of inserted CLASP-2 coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where the entire CLASP-2 gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals can be needed. However, in cases where only a portion of the CLASP-2 coding sequence is inserted, exogenous translational control signals, including the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the CLASP-2 coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, and the like. (see Bittner *et al.*, 1987, Methods in Enzymol. 153: 516-544).

In addition, a host cell strain can be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products can be important for the function of the protein. The presence of several consensus N-glycosylation sites in CLASP-2 extracellular domains support the possibility that proper modification can play a role in CLASP-2 function. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, WI38, and the like.

Host cells transformed with nucleotide sequences encoding CLASP-2 may be cultured under conditions suitable for the expression and recovery of the soluble protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CLASP-2 may be designed to contain signal sequences which direct secretion of CLASP-2 through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding CLASP-2 to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin,

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express CLASP-2 proteins can be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with the CLASP-2 DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, and the like.), and a selectable marker. Following the introduction of foreign DNA, engineered cells can be allowed to grow for 1-2 days in an enriched medium, and then switched to a selective medium. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines.

This method can advantageously be used to engineer cell lines which express the CLASP-2 protein(s) on the cell surface. Such engineered cell lines are particularly useful in screening for molecules or drugs that affect CLASP-2 function.

A number of selection systems can be used, including but not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977, Cell 11: 223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. U.S.A. 48: 2026), and adenine phosphoribosyltransferase (Lowy *et al.*, 1980, Cell 22: 817) genes which can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for *dhfr*, which confers resistance to methotrexate (Wigler *et al.*, 1980, Natl. Acad. Sci. U.S.A. 77: 3567; O'Hare *et al.*, 1981, Proc. Natl. Acad. Sci. U.S.A. 78: 1527); *gpt*, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981), Proc. Natl. Acad. Sci. U.S.A. 78: 2072); *neo*, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981, J. Mol. Biol. 150: 1); and *hygro*, which confers resistance to hygromycin (Santerre *et al.*, 1984, Gene 30: 147). Additional selectable genes have been described, namely *trpB*, which allows cells to utilize indole in place of tryptophan; *hisD*, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, 1988, Proc. Natl. Acad. Sci. U.S.A. 85: 8047); *ODC* (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue L., 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.) and glutamine synthetase (Bebbington *et al.*, 1992, Biotech 10: 169).

In an alternate embodiment of the invention, the coding sequence of CLASP-2 could be synthesized in whole or in part, using chemical methods well known in the art. (*See, e.g.,* Caruthers *et al.*, 1980, Nuc. Acids Res. Symp. Ser. 7: 215-233; Crea and Horn, 180, Nuc. Acids Res. 9(10): 2331; Matteucci and Caruthers, 1980, Tetrahedron Letter 21: 719; and Chow and Kempe, 1981, Nuc. Acids Res. 9(12): 2807-2817.) Alternatively, the protein itself could be produced using chemical methods to synthesize a CLASP-2 amino acid sequence in whole or in part. For example, peptides can be synthesized by solid phase techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography. (See Creighton, 1983, Proteins Structures And Molecular Principles, W.H. Freeman and Co., N.Y. pp. 50-60). The composition of the synthetic polypeptides can be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; see Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., N.Y., pp. 34-49).

In some embodiments, the CLASP-2 polypeptide contains non-naturally occurring amino acids or amino acid analogs (*i.e.*, compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium).

5.5.2. Identification of Cells That Express CLASP-2

The recombinant host cells which contain the coding sequence and which express a CLASP-2 gene product or fragments thereof can be identified by at least four general approaches; (a) DNA-DNA or DNA-RNA hybridization; (b) the presence or absence of "marker" gene functions; (c) assessing the level of transcription as measured by the expression of CLASP-2 mRNA transcripts in the host cell; and (d) detection of the gene product as measured by immunoassay or by its biological activity. Prior to the identification of gene expression, the host cells can be first mutagenized in an effort to increase the level of expression of CLASP-2, especially in cell lines that produce low amounts of CLASP-2.

In the first approach, the presence of the CLASP-2 coding sequence inserted in the expression vector can be detected by DNA-DNA or DNA-RNA hybridization using probes comprising nucleotide sequences that are homologous to the CLASP-2 coding sequence, respectively, or portions or derivatives thereof.

In the second approach, the recombinant expression vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, resistance to methotrexate, transformation phenotype, occlusion body formation in baculovirus, and the like). For example, if the CLASP-2 coding sequence is inserted within a marker gene sequence of the vector, recombinants containing the CLASP-2 coding sequence can be identified by the absence of the marker gene function. Alternatively, a marker gene can be placed in tandem with the CLASP-2 sequence under the control of the same or different promoter used to control the expression of the CLASP-2 coding sequence. Expression of the marker in response to induction or selection indicates expression of the CLASP-2 coding sequence.

In the third approach, transcriptional activity for the CLASP-2 coding region can be assessed by hybridization assays. For example, RNA can be isolated and analyzed by Northern blot using a probe homologous to the CLASP-2 coding sequence or particular portions thereof. Alternatively, total nucleic acids of the host cell can be extracted and

assayed for hybridization to such probes. Additionally, reverse transcription-polymerase chain reactions can be used to detect low levels of gene expression.

In the fourth approach, the expression of the CLASP-2 protein product can be assessed immunologically, for example by Western blots, immunoassays such as radioimmuno-precipitation, enzyme-linked immunoassays, fluorescent activated cell sorting ("FACS"), and the like. This can be achieved by using an anti-CLASP-2 antibody. Alternatively, CLASP-2 protein can be expressed as a fusion protein with green-fluorescent protein to facilitate its detection in cells (United States Patent Nos. 5,491,084; 5,804,387; 5,777,079).

Identification of cells or tissues expressing CLASP protein or mRNA, especially CLASP-2 isoforms, can be useful for determining normal and abnormal CLASP expression in a given cell or tissue. As discussed above, a number of CLASP-2 isoforms have been identified, *e.g.,* in Jurkat cells, peripheral blood, and brain. The identification of mRNA or protein expression in various cell types and tissues can allow for identification of isoforms improperly expressed in either a spatial or temporal manner. Expression of hCLASP-2D isoform in hematopoietic cells may cause problems due to the presence of the SH3 domain not seen in the Jurkat and peripheral blood isoforms.

Other molecules in the immune system may also interact with portions of hCLASP2D. However, the absence of the PBM domain in the hCLASP-2D isoform may be necessary for function in certain cell types or tissues. Similarly, expression of CLASP isoforms 2A, 2B, and 2C in brain may cause problems for different reasons: the PBM present in these isoforms may interfere with a particular function by binding any of the known PDZ domain protein involved in formation of the neurological synapse. Similarly, the lack of an SH3 domain may cause an inappropriate response due to interactions with only a subset of molecules required for CLASP-2 function in the brain.

5.5.3. Uses of CLASP-2 Engineered Host Cells

In one embodiment of the invention, the CLASP-2 protein and/or cell lines that express CLASP-2 can be used to screen for antibodies, peptides, small molecules, natural and synthetic compounds or other cell bound or soluble molecules that bind to the CLASP-2 protein resulting in stimulation or inhibition of CLASP-2 function. For example, anti-CLASP-2 antibodies can be used to inhibit or stimulate CLASP-2 function and to detect its presence. Alternatively, screening of peptide libraries with recombinantly expressed soluble CLASP-2 protein or cell lines expressing CLASP-2 protein can be useful for identification of

therapeutic molecules that function by inhibiting or stimulating the biological activity of CLASP-2. The uses of the CLASP-2 protein and engineered cell lines, described in the subsections below, can be employed equally well for homologous CLASP-2 genes in various species.

In a specific embodiment of the invention, cell lines may be engineered to express the extracellular or intracellular domain of CLASP fused to another molecule such as GST. In addition, CLASP, its extracellular domain or its intracellular domain may be fused to an immunoglobulin constant region (Hollenbaugh and Aruffo, 1992, Current Protocols in Immunology, Unit 10.19; Aruffo *et al.*, 1990, Cell 61: 1303) to produce a soluble molecule with increased half life. The soluble protein or fusion protein can be used in binding assays, affinity chromatography, immunoprecipitation, Western blot, and the like. Synthetic compounds, natural products, and other sources of potentially biologically active materials can be screened in assays that are well known in the art.

Random peptide libraries consisting of all possible combinations of amino acids attached to a solid phase support can be used to identify peptides that are able to bind to a specific domain of CLASP-2 (Lam, K.S. *et al.*, 1991, Nature 354: 82-84). The screening of peptide libraries can have therapeutic value in the discovery of pharmaceutical agents that stimulate or inhibit the biological activity of CLASP-2.

Identification of molecules that are able to bind to the CLASP-2 protein can be accomplished by screening a peptide library with recombinant soluble CLASP-2 protein. Methods for expression and purification of CLASP-2 are described in Section 5.7, *supra*, and can be used to express recombinant full length CLASP-2 or fragments of CLASP-2 depending on the functional domains of interest. Such domains include CLASP-2 extracellular domain, transmembrane domain, CLASP-2 intracellular domain, ITAM containing domain, tyrosine phosphorylation site containing domain, cysteine cluster containing domain, cadherin motif containing domain, and coil/coil domain.

To identify and isolate the peptide/solid phase support that interacts and forms a complex with CLASP-2, it is necessary to label or "tag" the CLASP-2 molecule. The CLASP-2 protein can be conjugated to enzymes such as alkaline phosphatase or horseradish peroxidase or to other reagents such as fluorescent labels which can include fluorescein isothiocyanate (FITC), phycoerythrin (PE) or rhodamine. Conjugation of any given label to CLASP-2 can be performed using techniques that are well known in the art. Alternatively, CLASP-2 expression vectors can be engineered to express a chimeric CLASP-2 protein containing an epitope for which a commercially available antibody exist. The epitope-

specific antibody can be tagged with a detectable label using methods well known in the art including an enzyme, a fluorescent dye or colored or magnetic beads.

The "tagged" CLASP-2 conjugate is incubated with the random peptide library for 30 minutes to one hour at 22°C to allow complex formation between CLASP-2 and peptide species within the library. The library is then washed to remove any unbound protein. If CLASP-2 has been conjugated to alkaline phosphatase or horseradish peroxidase the whole library is poured into a petri dish containing substrates for either alkaline phosphatase or peroxidase, for example, 5-bromo-4-chloro-3-indoyl phosphate (BCIP) or 3,3',4,4"-diaminobenzidine (DAB), respectively. After incubating for several minutes, the peptide/solid phase- CLASP-2 complex changes color, and can be easily identified and isolated physically under a dissecting microscope with a micromanipulator. If a fluorescent tagged CLASP-2 molecule has been used, complexes can be isolated by fluorescence activated sorting. If a chimeric CLASP-2 protein expressing a heterologous epitope has been used, detection of the peptide/CLASP-2 complex can be accomplished by using a labeled epitope-specific antibody. Once isolated, the identity of the peptide attached to the solid phase support can be determined by peptide sequencing.

In addition to using soluble CLASP-2 molecules, in another embodiment, it is possible to detect peptides that bind to cell-associated CLASP-2 using intact cells. The use of intact cells is preferred for use with cell surface molecules. Methods for generating cell lines expressing CLASP-2 are described in Section 5.8. The cells used in this technique can be either live or fixed cells. The cells can be incubated with the random peptide library and bind to certain peptides in the library to form a "rosette" between the target cells and the relevant solid phase support/peptide. The rosette can thereafter be isolated by differential centrifugation or removed physically under a dissecting microscope. Techniques for screening combinatorial libraries are known in the art (Gallop *et al.*, 1994, J. Med. Chem., 37: 1233; Gordon, 1994, J. Med. Chem., 37: 1385).

As an alternative to whole cell assays for membrane bound receptors or receptors that require the lipid domain of the cell membrane to be functional, CLASP-2 molecules can be reconstituted into liposomes where label or "tag" can be attached.

5.5.4. CLASP-2 Fusion Proteins

In another embodiment of the invention, a CLASP-2 or a modified CLASP-2 sequence can be ligated to a heterologous sequence to encode a fusion protein. For example, for screening of peptide libraries for molecules that bind CLASP-2, it can be useful to

produce a chimeric CLASP-2 protein expressing a heterologous epitope that is recognized by a commercially available antibody. A fusion protein can also be engineered to contain a cleavage site located between a CLASP-2 sequence and the heterologous protein sequence, so that the CLASP-2 can be cleaved away from the heterologous moiety. In one embodiment, fusion proteins of the invention can contain the CLASP-2 extracellular domain comprising at least about residues 1 through 816 or fragment thereof. In another embodiment, fusion proteins can contain the CLASP-2 intracellular domain comprising at least about residue 843 through the end of the CLASP-2 sequence or fragment thereof.

5.6. Cloning Alleles, Variants, and Species Homologs of CLASP-2

In order to clone the full length cDNA sequence from any species encoding a CLASP-2 cDNA, or to clone variant forms of the molecule, labeled DNA probes made from nucleic acid fragments corresponding to any partial cDNA disclosed herein can be used to screen a cDNA library derived from lymphoid cells or brain cells. More specifically, oligonucleotides corresponding to either the 5' or 3' terminus of the cDNA sequence can be used to obtain longer nucleotide sequences. Briefly, the library can be plated out to yield a maximum of 30,000 pfu for each 150 mm plate. Approximately 40 plates can be screened. The plates are incubated at 37°C until the plaques reach a diameter of 0.25 mm or are just beginning to make contact with one another (3-8 hours). Nylon filters are placed onto the soft top agarose and after 60 seconds, the filters are peeled off and floated on a DNA denaturing solution consisting of 0.4N sodium hydroxide. The filters are then immersed in neutralizing solution consisting of 1M Tris-HCl, pH 7.5, before being allowed to air dry. The filters are prehybridized in hybridization buffer such as casein buffer containing 10% dextran sulfate, 0.5M NaCl, 50mM Tris-HCl, pH 7.5, 0.1% sodium pyrophosphate, 1% casein, 1% SDS, and denatured salmon sperm DNA at 0.5 mg/ml for 6 hours at 60°C. The radiolabeled probe is then denatured by heating to 95°C for 2 minutes and then added to the prehybridization solution containing the filters. The filters are hybridized at 60°C for 16 hours. The filters are then washed in 1X wash mix (10X wash mix contains 3M NaCl, 0.6M Tris base, and 0.02M EDTA) twice for 5 minutes each at room temperature, then in 1X wash mix containing 1% SDS at 60°C for 30 minutes, and finally in 0.3X wash mix containing 0.1% SDS at 60°C for 30 minutes. The filters are then air dried and exposed to x-ray film for autoradiography. After developing, the film is aligned with the filters to select a positive plaque. If a single, isolated positive plaque cannot be obtained, the agar plug containing the

plaques will be removed and placed in lambda dilution buffer containing 0.1M NaCl, 0.01M magnesium sulfate, 0.035M Tris HCl, pH 7.5, 0.01% gelatin. The phage can then be replated and rescreened to obtain single, well isolated positive plaques. Positive plaques can be isolated and the cDNA clones sequenced using primers based on the known cDNA sequence. This step can be repeated until a full length cDNA is obtained.

It can be necessary to screen multiple cDNA libraries from different tissues to obtain a full length cDNA. In the event that it is difficult to identify cDNA clones encoding the complete 5' terminal coding region, an often encountered situation in cDNA cloning, the RACE (Rapid Amplification of cDNA Ends) technique can be used. RACE is a proven PCR-based strategy for amplifying the 5' end of incomplete cDNAs. 5'-RACE-Ready RNA synthesized from human tissues containing a unique anchor sequence is commercially available (Clontech). To obtain the 5' end of the cDNA, PCR is carried out on 5'-RACE-Ready cDNA using the provided anchor primer and the 3' primer. A secondary PCR reaction is then carried out using the anchored primer and a nested 3' primer according to the manufacturer's instructions. Once obtained, the full length cDNA sequence can be translated into amino acid sequence and examined for certain landmarks such as a continuous open reading frame flanked by translation initiation and termination sites, a cadherin-like domain, an ITAM domain, a tyrosine phosphorylation site, a cysteine cluster, a transmembrane domain, and finally overall structural similarity to the CLASP-2 genes disclosed herein. *See*, Ponassi *et al.*, 1999, Mech. Dev. 80: 207-212; Isakov, 1998, Receptor Channels 5: 243-253; Borroto *et al.*, 1997, Biopolymers 42: 75-88; Dimitratos *et al.*, 1997, Mech. Dev. 63: 127-130; Apperson *et al.*, 1996, J. Neurosci. 16: 6839-6852; Ozawa *et al.*, 1990, Mech. Dev. 33: 49-56, which discuss protein domains and are incorporated herein by reference.

5.7. Modulating Expression of Endogenous CLASP-2 Genes

Alternatively, the expression characteristics of an endogenous CLASP-2 gene within a cell population can be modified by inserting a heterologous DNA regulatory element into the genome of the cell line such that the inserted regulatory element is operatively linked with the endogenous CLASP-2 gene. For example, an endogenous CLASP-2 gene which is normally "transcriptionally silent", *i.e.*, an *CLASP-2* gene which is normally not expressed, or is expressed only at very low levels in a cell population, can be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in the cells. Alternatively, a transcriptionally silent, endogenous CLASP-2 gene

can be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element can be inserted into a cell line population, such that it is operatively linked with an endogenous CLASP-2 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, (see *e.g.*. in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published Jan 16, 1991).

5.8. Anti-CLASP-2 Antibodies

Various procedures known in the art can be used for the production of antibodies to epitopes of the natural and recombinantly produced CLASP-2 protein. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, human or humanized, IgG, IgM, IgA, IgD or IgE, a complementarity determining region, Fab fragments, $F(ab')_2$ and fragments produced by an Fab expression library as well as anti-idiotypic antibodies. Antibodies which compete for CLASP-2 binding are especially preferred for diagnostics and therapeutics.

Monoclonal antibodies that bind CLASP-2 can be radioactively labeled allowing one to follow their location and distribution in the body after injection. Radioisotope tagged antibodies can be used as a non-invasive diagnostic tool for imaging *de novo* lymphoid tumors and metastases that express CLASP-2.

Immunotoxins can also be designed which target cytotoxic agents to specific sites in the body. For example, high affinity CLASP-2 specific monoclonal antibodies can be covalently complexed to bacterial or plant toxins, such as diphtheria toxin or ricin. A general method of preparation of antibody/hybrid molecules can involve use of thiol-crosslinking reagents such as SPDP, which attack the primary amino groups on the antibody and by disulfide exchange, attach the toxin to the antibody. The hybrid antibodies can be used to specifically eliminate CLASP-2 expressing lymphocytes.

For the production of antibodies, various host animals can be immunized by injection with the recombinant or naturally purified CLASP-2 protein, fusion protein or peptides, including but not limited to goats, rabbits, mice, rats, hamsters, and the like Various adjuvants can be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and poten-

tially useful human adjuvants such as BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum*.

Monoclonal antibodies to CLASP-2 can be prepared by using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique originally described by Kohler and Milstein, (*Nature,* 1975, 256: 495-497), the human B-cell hybridoma technique (Kosbor *et al.,* 1983, Immunology Today, 4: 72; Cote *et al.,* 1983, Proc. Natl. Acad. Sci. U.S.A., 80: 2026-2030) and the EBV-hybridoma technique (Cole *et al.,* 1985, Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.,* 1984, Proc. Natl. Acad. Sci. U.S.A., 81: 6851-6855; Neuberger *et al.*, 1984, Nature, 312: 604-608; Takeda *et al.*, 1985, Nature, 314: 452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce CLASP-2 -specific single chain antibodies. In some embodiments, phage display technology is used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (see, *e.g.,* McCafferty *et al.*, Nature 348: 552-554 (1990); Marks *et al.*, Biotechnology 10: 779-783 (1992)).

Hybridomas can be screened using enzyme-linked immunosorbent assays (ELISA) in order to detect cultures secreting antibodies specific for refolded recombinant CLASP-2. Cultures can also be screened by ELISA to identify those cultures secreting antibodies specific for mammalian-produced CLASP-2. Confirmation of antibody specificity can be obtained by western blot using the same antigens. Subsequent ELISA testing can use recombinant CLASP-2 fragments to identify the specific portion of the CLASP-2 molecule with which a monoclonal antibody binds. Additional testing can be used to identify monoclonal antibodies with desired functional characteristics such as staining of histological sections, immunoprecipitation of CLASP-2, inhibition of CLASP-2 binding or stimulation of CLASP-2 to transmit an intracellular signal. Determination of the monoclonal antibody isotype can be accomplished by ELISA, thus providing additional information concerning purification or function.

Some anti-CLASP-2 monoclonal antibodies of the present invention are humanized, human or chimeric, in order to reduce their potential antigenicity, without reducing their affinity for their target. Humanized antibodies have been described in the art.

See, *e.g.*, Queen, *et al.*, 1989, Proc. Natl Acad. Sci. U.S.A. 86: 10029; U.S. Patent Nos. 5,563,762; 5,693,761; 5,585,089 and 5,530,101. The human antibody sequences used for humanization can be the sequences of naturally occurring human antibodies or can be consensus sequences of several human antibodies. See Kettleborough *et al.*, 1991, *Protein Engineering* 4: 773; Kolbinger *et al.*, 1993, Protein Engineering 6: 971. Humanized monoclonal antibodies against CLASP-2 peptides can also be produced using transgenic animals having elements of a human immune system (see, *e.g.*, U.S. Patent Nos. 5,569,825; 5,545,806; 5,693,762; 5,693,761; and 5,7124,350).

In some embodiments, an anti-CLASP-2 polypeptide monoclonal or polyclonal antiserum is produced that is specifically immunoreactive with a particular CLASP-2 polypeptide and is selected to have low cross-reactivity against other molecules (*e.g.*, other CLASP polypeptides) and any such cross-reactivity is removed by immunoabsorbtion prior to use in the immunoassay. Methods for screening and characterizing monoclonal antibodies for specificity are well known in the art and are described generally in Harlow and Lane, *supra*. For example, polyclonal antibodies raised to hCLASP-2A, as shown in SEQ ID NO: 1, or splice variants, or immunogenic portions thereof, can be selected to obtain only those polyclonal or monoclonal antibodies that are specifically immunoreactive with the target protein not with other proteins. This selection may be achieved by subtracting out antibodies that cross-react with molecules. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, *e.g.*, Harlow & Lane, Antibodies, A Laboratory Manual (1988) for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background. Alternatively, antibodies that cross-react with a selected set of polypeptides may be prepared.

Antibody fragments which contain specific binding sites of V can be generated by known techniques. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed (Huse *et al.*, 1989, Science, 246: 1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity to CLASP-2.

Anti-CLASP-2 antibodies can also be used to identify, isolate, inhibit or eliminate CLASP-2-expressing cells. In one embodiment, the present invention includes a method of identifying an abnormal T cell profile of an immunocompromised subject relative to the T cell profile of a non-immunocompromised subject. The method includes (i) sorting a sample of peripheral blood mononuclear cells (PBMC) isolated from the immunocompromised subject into sets of T cell types, (ii) determining the ratio of CLASP-2^+ cells relative to the total number of cells (CLASP-2^+: total) in each set, and identifying an abnormal T cell profile in the immunocompromised subject by comparing the CLASP-2^+: total ratios of sets from the immunocompromised subject with the CLASP-2^+: total ratios of analogous sets from a non-immunocompromised subject.

In other embodiments, anti-CLASP-2 antibodies can be used for detection of hCLASP-2 protein in assays such as fluorescent activated cell sorting (FACS), ELISA, fluorescent or electron immunomicroscopy, Western blots, gel shift analyses. CLASP-2 expression in various cells, localization within cells, interactions with other proteins, and differentiation between CLASP-2 isoform expression can be determined by use of the techniques listed herein.

5.9. Screening Assays

The invention provides methods for identifying compounds or agents that modulate (*i.e.*, inhibit or enhance) CLASP-2 expression or activity. CLASP-2 expression or activity modulators are useful for treatment of disorders characterized by (or associated with) aberrant or abnormal CLASP-2 expression or activity. Aberrant expression of CLASP-2 mRNA or protein means expression in lymphocytes (*e.g.*, T lymphocytes or B lymphocytes) or other CLASP-2 expressing cells of at least 2-fold, preferably at least 5-fold greater than expression in control lymphocytes obtained from a healthy subject.

The CLASP-2 expression assays can include the steps of contacting a cell expressing CLASP-2 with a compound or agent and assaying CLASP-2 expression. CLASP-2 polypeptide expression is easily measured by ELISA using anti-CLASP-2 antibodies of the invention. CLASP-2 mRNA expression (including expression of specific species or splice variants of CLASP-2) can be measured by quantitative Northern analysis or quantitative PCR.

CLASP-2 activities include, for example, the CLASP-2 polypeptide binding to PDZ-domain containing molecules and CLASP-2 polypeptide involvement in signal transduction (*e.g.*, leading to T cell activation). Compounds or agents that modulate the

interaction of a CLASP-2 polypeptide and a target molecule, modulate CLASP-2 nucleic acid expression, or modulate CLASP-2 polypeptide activity are all contemplated by the methods of the present invention.

Test compounds include, for example, 1) peptides (*e.g.*, soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, *e.g.*, Lam, K. S. *et al.*, 1991, Nature 354: 82-84; Houghten, R. *et al.*, 1991, Nature 354: 84-86) and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids; 2) phosphopeptides (*e.g.*, members of random and partially degenerate, directed phosphopeptide libraries, see, *e.g.*, Songyang, Z. *et al.*, 1993, Cell 72: 767-778); 3) CLASP-2 antibodies (as described above); 4) small organic and inorganic molecules (*e.g.*, molecules obtained from combinatorial and natural product libraries); 5) antisense RNA and DNA molecules and ribozymes (described above).

The CLASP modulators can be any of a large variety of compounds, both naturally occurring and synthetic, organic and inorganic, and including polymers (e.g., oligopeptides, polypeptides, oligonucleotides, and polynucleotides), small molecules, antibodies, sugars, fatty acids, nucleotides and nucleotide analogs, analogs of naturally occurring structures (e.g., peptide mimetics, nucleic acid analogs, and the like), and numerous other compounds.

In one embodiment, the invention provides assays for screening test compounds which bind to CLASP-2 polypeptides. The assays can be recombinant cell based or cell-free assays. These assays can include the steps of combining a cell expressing a CLASP-2 polypeptide or a binding fragment thereof, and a compound or agent under conditions which allow binding of the compound or agent to the CLASP-2 polypeptide to form a complex. Complex formation can then be determined. The ability of the candidate compound or agent to bind to the CLASP-2 polypeptide or fragment thereof is indicated by the presence of the candidate compound in the complex. Formation of complexes between the CLASP-2 polypeptide and the candidate compound can be quantitated, for example, using standard immunoassays.

In another embodiment, the invention provides screening assays to identify test compounds which modulate the interaction (and most likely CLASP-2 activity as well) between a CLASP-2 polypeptide and a molecule (target molecule with which the CLASP-2 polypeptide normally interacts.

In one embodiment, these CLASP-2 target molecules can be tyrosine kinases (*e.g.*, lyn, lck, fyn, ZAP-70m SyK, and CSK). In another embodiment, these CLASP-2 target

molecules can be tyrosine phosphatases (*e.g.*, EZRIN, SHP-1, SHP-2 and PTP36). In another embodiment, these CLASP-2 target molecules can be adaptor proteins (*e.g.*, NCK, CBL, SHC, LNK, SLP-76, HS1, SIT, VAV, GrB2, and BRDG1). In another embodiment, these CLASP-2 target molecules can be cytoskeletal associated proteins such as ankyrin, spectrin, talin, ezrin, tropomyosin, myosin, plectin, syndecans, paralemmin, Band 3 protein, cytoskeletal protein 4.1, and PTP36. In a further embodiment, CLASP-2 target molecules can be members of the integrin family.

Typically, the assays are recombinant cell based or cell-free assay. These assays can include the steps of combining a cell expressing a CLASP-2 polypeptide or a binding fragment thereof, a CLASP-2 target molecule (*e.g.*, a CLASP-2 ligand) and a test compound, under conditions where but for the presence of the candidate compound, the CLASP-2 polypeptide or biologically active portion thereof binds to the target molecule. Detecting complex formation between the CLASP-2 polypeptide or the binding fragment thereof, the CLASP-2 target molecule and a test compound detecting the formation of a complex which includes the CLASP-2 polypeptide and the target molecule can be accomplished. Detection of complex formation can include direct quantitation of the complex by, for example, measuring inductive effects, such as T cell activation, of the CLASP-2 polypeptide. A significant change, such as a decrease, in the interaction of the CLASP-2 and target molecule (*e.g.*, in the formation of a complex between the CLASP-2 and the target molecule) in the presence of a candidate compound (relative to what is detected in the absence of the candidate compound) is indicative of a modulation of the interaction between the CLASP-2 polypeptide and the target molecule. Modulation of the formation of complexes between the CLASP-2 polypeptide and the target molecule can be quantitated using, for example, an immunoassay. To perform cell free drug screening assays, it is desirable to immobilize either CLASP-2 or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. CLASP-2 binding to a target molecule, in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and microcentrifuge tubes.

In one embodiment, a fusion polypeptide can be provided which adds a domain that allows the polypeptide to be bound to a matrix. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of CLASP-2-binding

polypeptide found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing polypeptides on matrices can also be used in the drug screening assays of the invention. For example, either CLASP-2 or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated CLASP-2 molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with CLASP-2 but which do not interfere with binding of the polypeptide to its target molecule can be derivatized to the wells of the plate, and CLASP-2 trapped in the wells by antibody conjugation. As described above, preparations of a CLASP-2 -binding polypeptide and a candidate compound are incubated in the CLASP-2 -presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes include immunodetection of complexes using antibodies reactive with the CLASP-2 target molecule, or which are reactive with CLASP-2 polypeptide and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant DNA molecules expressing the CLASP-2, *e.g.*, the protein having the sequence of SEQ ID NO: 2. Such cells, either in viable or fixed form, can be used for standard ligand/receptor binding assays (see, *e.g.*, Parce *et al.* (1989) Science 246: 243-247; and Owicki *et al.* (1990) Proc. Natl Acad. Sci. U.S.A. 87: 4007-4011, which describe sensitive methods to detect cellular responses. A test compound, often labeled, can be assayed for binding or for competition with another ligand for binding. Viable cells could also be used to screen for the effects of drugs on CLASP-2 mediated functions, *e.g.*, T cell activation, second messenger levels, and others).

In another embodiment, the invention provides a method for identifying a compound (*e.g.*, a screening assay) capable of use in the treatment of a disorder characterized by (or associated with) aberrant or abnormal CLASP-2 nucleic acid expression or CLASP-2 polypeptide activity. This method typically includes the step of assaying the ability of the compound or agent to modulate the expression of the CLASP-2 nucleic acid or the activity of the CLASP-2 polypeptide thereby identifying a compound for treating a disorder characterized by aberrant or abnormal CLASP-2 nucleic acid expression or CLASP-2 polypeptide activity.

Methods for assaying the ability of the compound or agent to modulate the expression of the CLASP-2 nucleic acid or activity of the CLASP-2 polypeptide are typically cell-based assays. For example, cells which are sensitive to ligands which transduce signals via a pathway involving CLASP-2 can be induced to overexpress a CLASP-2 polypeptide in the presence and absence of a candidate compound. Candidate compounds which produce a change in CLASP-2-dependent responses can be identified. In one embodiment, expression of the CLASP-2 nucleic acid or activity of a CLASP-2 polypeptide is modulated in cells and the effects of candidate compounds on the readout of interest (such as T cell activation) are measured. For example, the expression of genes which are up- or down-regulated in response to a CLASP-2-dependent signal cascade can be assayed.

Alternatively, modulators of CLASP-2 expression can be identified in a method where a cell is contacted with a candidate compound and the expression of CLASP-2 mRNA or polypeptide in the cell is determined. The level of expression of CLASP-2 mRNA or polypeptide in the presence of the candidate compound is compared to the level of expression of CLASP-2 mRNA or polypeptide in the absence of the candidate compound. The candidate compound can then be identified as a modulator of CLASP-2 nucleic acid expression based on this comparison. For example, when expression of CLASP-2 mRNA or polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of CLASP-2 nucleic acid expression. Alternatively, when CLASP-2 nucleic acid expression is less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of CLASP-2 nucleic acid expression. The level of CLASP-2 nucleic acid expression in the cells can be determined by methods described herein for detecting CLASP-2 mRNA or polypeptide.

Modulators of CLASP-2 polypeptide activity and CLASP-2 nucleic acid expression identified according to these drug screening assays can be used to treat, for example, immune disorders. These methods of treatment include the steps of administering the modulators of CLASP-2 polypeptide activity or nucleic acid expression, *e.g.*, in a pharmaceutical composition as described in §5.10.1 below, to a subject in need of such treatment, *e.g.*, a subject with a disorder described herein.

5.10. Therapeutic Administration of CLASP-2 Modulators

The CLASP-2 protein is expressed in lymphocytes and, as noted *supra*, play a role in regulating T cell and B cell interactions, thus making CLASP-2 activity (*e.g.*, CLASP-

2 binding of regulatory proteins) a target for diagnostic and treatment of immune disorders and for modulation of immune function (*e.g.*, T cell activation). Additionally, since CLASP-2 contains domains capable of transducing an intracellular signal, cell surface CLASP-2 can be triggercd by an anti- CLASP-2 antibody or soluble CLASP-2 or a fragment thereof in order to enhance thc activation state of a lymphocyte.

5.10.1. Formulation and Route of Administration

A CLASP-2 polypeptide, a fragment thereof, anti-CLASP-2 antibody, CLASP-2 polynucleotide (*e.g.*, antisense or ribozyme), or small molecule agonists or antagonists can be administered to a subject *per se* or in the form of a pharmaceutical or therapeutic composition. Pharmaceutical compositions comprising the proteins of the invention can be manufactured by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Pharmaceutical compositions can be formulated in conventional manner using one or more physiologically acceptable carriers, diluents, excipients or auxiliaries which facilitate processing of the protein or active peptides into preparations which can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

Currently, there are three major classes of protein-derived cell-penetrating peptides that have been used for delivering of proteins into cells and animals (Lindgren, M.; *et al.*, 2000, Trcnds Pharmacol Sci. 21: 99-103). In one embodiment, the CLASP-2 protein or fragment (encoding a functional domain of CLASP-2) can be introduced into the cell as a fusion protein tied to a transporter protein derived from homeoprotein transcription factors such as ANTP. In another embodiment, the CLASP-2 protein or fragment (encoding a functional domain of CLASP-2) can be introduced into the cell as a fusion protein tied to other transcription factors such as the HIV Tat protein and the herpes simplex virus type 1 (HSV-1) VP22 protein. Members in this family have been widely used in different cellular and animal systems (Schwarze, S.; *et al.*; 2000, Trends Pharmacol Sci. 21: 45-48). In another embodiment, the CLASP-2 protein or fragment (encoding a functional domain of CLASP-2) can be introduced into the cell as a fusion protein tied to peptides derived from signal-sequences present in several proteins such as HIV-1 gp41. In other embodiments, there are several synthetic and/or chemeric cell-penetrating peptides such as transportan and Amphiphiloc model peptide (Lindgren, M.; *et al.*, 2000, Trends Pharmacol Sci. 21: 99-103) that can be used. In another embodiment, the CLASP-2 protein or fragment can be

introduced by using anti-DNA antibodies (see, *e.g.*, Zack, D. J., *et al.*, 1996, J. Immunol. 157: 2082-8

For topical administration the proteins of the invention can be formulated as solutions, gels, ointments, creams, suspensions, and the like. as are well-known in the art.

Systemic formulations include those designed for administration by injection, *e.g.*, subcutaneous, intravenous, intramuscular, intrathecal or intraperitoneal injection, as well as those designed for transdermal, transmucosal, oral or pulmonary administration.

For injection, the proteins of the invention can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. The solution can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the proteins can be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, a composition can be readily formulated by combining the proteins with pharmaceutically acceptable carriers well known in the art. Such carriers enable the proteins to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. For oral solid formulations such as, for example, powders, capsules and tablets, suitable excipients include fillers such as sugars, such as lactose, sucrose, mannitol and sorbitol; cellulose preparations such as maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP); granulating agents; and binding agents. If desired, disintegrating agents can be added, such as the cross-linked polyvinylpyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

If desired, solid dosage forms can be sugar-coated or enteric-coated using standard techniques.

For oral liquid preparations such as, for example, suspensions, elixirs and solutions, suitable carriers, excipients or diluents include water, glycols, oils, alcohols, and the like. Additionally, flavoring agents, preservatives, coloring agents and the like can be added.

For buccal administration, the proteins can take the form of tablets, lozenges, and the like. formulated in conventional manner.

For administration by inhalation, the proteins for use according to the present invention are conveniently delivered in the form of an aerosol spray from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The proteins can also be formulated in rectal or vaginal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the proteins can also be formulated as a depot preparation. Such long acting formulations can be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the proteins can be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Alternatively, other pharmaceutical delivery systems can be employed. Liposomes and emulsions are well known examples of delivery vehicles that can be used to deliver the proteins or peptides of the invention. Certain organic solvents such as dimethylsulfoxide also can be employed, although usually at the cost of greater toxicity. Additionally, the proteins can be delivered using a sustained-release system, such as semipermeable matrices of solid polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules can, depending on their chemical nature, release the proteins for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization can be employed.

As the proteins and peptides of the invention can contain charged side chains or termini, they can be included in any of the above-described formulations as the free acids or bases or as pharmaceutically acceptable salts. Pharmaceutically acceptable salts are those salts which substantially retain the biologic activity of the free bases and which are prepared by reaction with inorganic acids. Pharmaceutical salts tend to be more soluble in aqueous and other protic solvents than are the corresponding free base forms.

5.10.2. Effective Dosages

CLASP-2 polypeptides, CLASP-2 fragments and anti-CLASP-2 antibodies will generally be used in an amount effective to achieve the intended purpose. For use to inhibit an immune response, the proteins of the invention, or pharmaceutical compositions thereof, are administered or applied in a therapeutically effective amount. By therapeutically effective amount is meant an amount effective ameliorate or prevent the symptoms, or prolong the survival of, the patient being treated. Determination of a therapeutically effective amount is well within the capabilities of those skilled in the art, especially in light of the detailed disclosure provided herein.

For systemic administration, a therapeutically effective dose can be estimated initially from *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of test compound that inhibits 50% of CLASP-2 binding interactions). Such information can be used to more accurately determine useful doses in humans.

Initial dosages can also be estimated from *in vivo* data, *e.g.*, animal models, using techniques that are well known in the art. One having ordinary skill in the art could readily optimize administration to humans based on animal data.

Dosage amount and interval can be adjusted individually to provide plasma levels of the proteins which are sufficient to maintain therapeutic effect. Usual patient dosages for administration by injection range from about 0.1 to 5 mg/kg/day, preferably from about 0.5 to 1 mg/kg/day. Therapeutically effective serum levels can be achieved by administering multiple doses each day.

In cases of local administration or selective uptake, the effective local concentration of the proteins can not be related to plasma concentration. One having skill in the art will be able to optimize therapeutically effective local dosages without undue experimentation.

The amount of CLASP-2 administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

The therapy can be repeated intermittently while symptoms detectable or even when they are not detectable. The therapy can be provided alone or in combination with other drugs. In the case of autoimmune disorders, the drugs that can be used in combination

with CLASP-2 or fragments thereof include, but are not limited to, steroid and non-steroid immunosuppressive agents.

5.10.3. Toxicity

Preferably, a therapeutically effective dose of the proteins described herein will provide therapeutic benefit without causing substantial toxicity.

Toxicity of the proteins described herein can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, by determining the LD_{50} (the dose lethal to 50% of the population) or the LD_{100} (the dose lethal to 100% of the population). The dose ratio between toxic and therapeutic effect is the therapeutic index. The data obtained from these cell culture assays and animal studies can be used in formulating a dosage range that is not toxic for use in human. The dosage of the proteins described herein lies preferably within a range of circulating concentrations that include the effective dose with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (*See, e.g.*, Fingl *et al.*, 1975, In: The Pharmacological Basis of Therapeutics, Ch.1, p.1).

5.11. Binding Assays

CLASP-2 polypeptides can be used to screen for molecules that bind to CLASP-2 or for molecules to which CLASP-2 binds. The binding of CLASP-2 by the molecule can activate (agonist), increase, inhibit (antagonist), or decrease activity of the CLASP-2 or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (*e.g.*, receptors), or small molecules. Preferably, the molecule is closely related to the natural ligand of CLASP-2, *e.g.*, a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan *et al.*, Current Protocols in Immunology 1(2): Chapter 5 (1991).) Similarly, the molecule can be closely-related to the natural receptor to which CLASP-2 binds, or at least, a fragment of the receptor capable of being bound by CLASP-2 (*e.g.*, active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express CLASP-2, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing CLASP-2

(or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either CLASP-2 or the molecule.

The assay can simply test binding of a candidate compound to CLASP-2, where binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay can test whether the candidate compound results in a signal generated by binding to CLASP-2.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide affixed to a solid support, chemical libraries, or natural product mixtures. The assay can also simply comprise the steps of mixing a candidate compound with a solution containing CLASP-2, measuring CLASP-2 activity or binding, and comparing the CLASP-2 activity or binding to a standard. Preferably, an ELISA assay can measure CLASP-2 level or activity in a sample (*e.g.*, biological sample) using a monoclonal or polyclonal antibody. The antibody can measure CLASP-2 level or activity by either binding, directly or indirectly, to CLASP-2 or by competing with CLASP-2 for a substrate.

In another aspect of the invention, the CLASP-2 polypeptides, or fragments thereof, can be used as "bait proteins" in a two-hybrid assay (see, *e.g.*, U.S. Pat. No. 5,283,317; Zervos *et al.*, 1993, Cell 72: 223-232; Madura *et al.*, 1993, J. Biol. Chem. 268: 12046-12054; Bartel *et al.*, 1993, Biotechniques 14: 920-924; Iwabuchi *et al.*, 1993, Oncogene 8: 1693-1696; and Brent WO 94/10300), to identify other proteins, which bind to or interact with CLASP-2 ("CLASP-2-binding proteins" or "CLASP-2-bp") and modulate CLASP-2 polypeptide activity. Such CLASP-2-binding proteins are also likely to be involved in the propagation of signals by the CLASP-2 polypeptides as, for example, upstream or downstream elements of the CLASP-2 pathway.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient by activating or inhibiting the CLASP-2 molecule. Moreover, the assays can discover agents which can inhibit or enhance the production of CLASP-2 from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds or agents that bind to CLASP-2 polypeptides comprising the steps of: (a) contacting a CLASP-2 polypeptide with a compound or agent under conditions which allow binding of the compound to the CLASP-2 polypeptide to form a complex and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists or

antagonists comprising the steps of: (a) incubating a candidate compound with CLASP-2, (b) assaying a biological activity, and (b) determining if a biological activity of CLASP-2 has been altered.

Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds in a short period. See, *e.g.*, Fodor *et al.*, 1991, Science 251: 767-773, and other descriptions of chemical diversity libraries, which describe means for testing of binding affinity by a plurality of compounds.

5.12. Other Uses of CLASP-2 Polynucleotides and Polypeptides

The polynucleotides, polypeptides, polypeptide homologues, modulators, and antibodies described herein can be used in one or more of the following methods: a) drug screening assays; b) diagnostic assays particularly in disease identification, allelic screening and pharmocogenetic testing; and c) pharmacogenomics. A CLASP-2 polypeptide of the invention has one or more of the activities described herein and can thus be used to, for example, modulate an immune response in an immune cell, for example by binding to a CLASP-2 binding partner making it unavailable for binding to the naturally present CLASP-2 polypeptide.

In one embodiment, these CLASP-2 binding partners can be tyrosine kinases (*e.g.*, lyn, lck, fyn, ZAP-70m SyK, and CSK). In another embodiment, these CLASP-2 binding partners can be tyrosine phosphatases (*e.g.*, EZRIN, SHP-1, SHP-2 and PTP36). In another embodiment, these CLASP-2 target molecules can be adaptor proteins (*e.g.*, NCK, CBL, SHC, LNK, SLP-76, HS1, SIT, VAV, GrB2, and BRDG1. In another embodiment, these CLASP-2 binding partners can be cytoskeletal associated proteins such as ankyrin, spectrin, talin, ezrin, tropomyosin, myosin, plectin, syndecans, paralemmin, Band 3 protein, cytoskeletal protein 4.1, and PTP36. In a further embodiment, CLASP-2 binding partners can be members of the integrin family.The isolated nucleic acid molecules of the invention can be used to express CLASP-2 polypeptide (*e.g.*, via a recombinant expression vector in a host cell or in gene therapy applications), to detect CLASP-2 mRNA (*e.g.*, in a biological sample) or a naturally occurring or recombinantly generated genetic mutation in an CLASP-2 gene, and to modulate CLASP-2 activity, as described further below. In addition, the CLASP-2 polypeptides can be used to screen drugs or compounds which modulate CLASP-2 polypeptide activity as well as to treat disorders characterized by insufficient production of CLASP-2 polypeptide or production of CLASP-2 polypeptide forms which have decreased activity compared to wild type CLASP-2. Moreover, the anti-CLASP-2 antibodies of the

invention can be used to detect and isolate an CLASP-2 polypeptide, particularly fragments of CLASP-2 present in a biological sample, and to modulate CLASP-2 polypeptide activity.

5.13. Diagnostic Assays

The invention further provides a method for detecting the presence of CLASP-2, or fragment thereof, in a biological sample. Usually the biological sample contains lymphocytes (*e.g.*, from blood). The method involves contacting the biological sample with a compound or an agent capable of detecting CLASP-2 polypeptide or mRNA such that the presence of CLASP-2 is detected in the biological sample.

A preferred agent for detecting CLASP-2 mRNA is a directly or indirectly labeled nucleic acid probe capable of hybridizing to CLASP-2 mRNA. The nucleic acid probe can be, for example, the full-length CLASP-2 cDNA of SEQ ID NO: 1, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to CLASP-2 mRNA.

A preferred agent for detecting CLASP-2 polypeptide is a directly or indirectly labeled antibody capable of binding to a CLASP-2 polypeptide. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab)_2$) can be used. The term "directly or indirectly", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The detection method of the invention can be used to detect CLASP-2 mRNA or polypeptide in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of CLASP-2 mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of CLASP-2 polypeptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, CLASP-2 polypeptide can be detected in vivo in a subject by introducing into the subject a labeled anti-CLASP-2 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of CLASP-2 expressed in a subject and methods which detect fragments of an CLASP-2 polypeptide in a sample.

The invention also encompasses kits for detecting the presence of CLASP-2 in a biological sample. For example, the kit can comprise a directly or indirectly labeled compound or agent capable of detecting CLASP-2 polypeptide or mRNA in a biological sample; means for determining the amount of CLASP-2 in the sample; and means for comparing the amount of CLASP-2 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect CLASP-2 mRNA or polypeptide.

The methods of the invention can also be used to detect naturally occurring genetic mutations in an CLASP-2 gene, thereby determining if a subject with the mutated gene is at risk for a disorder characterized by aberrant or abnormal CLASP-2 nucleic acid expression or CLASP-2 polypeptide activity as described herein. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic mutation characterized by at least one of an alteration affecting the integrity of a gene encoding an CLASP-2 polypeptide, or the misexpression of the CLASP-2 gene.

5.14. Biological Activities of CLASP-2

As described herein, CLASP-2 mediates a variety of cell functions in lymphocytes and other cells. As described herein, a variety of assays are useful for detecting or quantitating CLASP-2 activity, or for identifying agents (including polynucleotides, polypeptides, and antibodies of the invention) that modulate CLASP-2 activity (*i.e.*, biological activity, *e.g.*, binding) or expression. Such agents are useful for treatment of diseases and conditions associated with aberrant CLASP-2 expression or activity. Further, following the guidance provided herein, other CLASP-2-mediated activities can be identified by those of skill using routine assays, such as those described below.

Exemplary assays for CLASP-2 function (or modulation of function) include assays for modulation of an *in vitro* or *in vivo* cell response (*e.g.*, an immune response such as lymphocyte activation, antibody production, inflammation) by detecting a change in an activity (*e.g.*, cytokine production, calcium flux, tyrosine phosphorylation, regulation of early activation markers, cell metabolism, proliferation, and the like, as described below) of cells *in vitro* or *in vivo*. In one embodiment, the cells are lymphocytes.

In one assay, for example, recombinant CLASP-2 protein, peptides, or antibodies corresponding to the CLASP-2 extracellular domain can be mixed directly with T and B cells. Cytokine production by these cells can then be measured and the degree of modulation of the immune response quantitated. Alternatively, antigen-presenting B cells are

mixed with untransfected T cells or T cells that have been transfected with CLASP-2 isoforms. Cytokine production (or calcium flux or other assays in §5.14.3) is be measured at the appropriate time to determine the effect of CLASP-2 on such an immune response. In a similar assay, B cells transfected with CLASP-2 constructs are tested for their ability to stimulate a T cell to generate an immune response. Transfected constructs in any of these cases could encode, for example, full or partial length CLASP-2 sequences, or antisense constructs to inhibit translation of endogenous CLASP-2 gene. Any of the examples described herein can be used to stimulate an immune response in the presence or absence of CLASP-2 isoforms or antibodies and assay the resulting effects on immune response by the methods listed in §5.14.3.

5.14.1 Methods for Generating an Immune Response in vitro

In various assays, an effect of an agent on immune cells is detected using an *in vitro* assay. The degree of an immune response can be measured or quantitated by a number of standard assays including those described below.

In one assay, human peripheral blood mononuclear cells (PBMC), human T cell clones (*e.g.*, Jurkat E6, ATCC TIB-152), EBV-transformed B cell clones (*e.g.*, 9D10, ATCC CRL-8752), antigen-specific T cell clones or lines can be used to examine immune responses in vitro. Activation, enhanced activation or inhibition of activation of these cells or cell lines can be used for the evaluation of potential CLASP therapeutics. Standard methods by which hematopoietic cells are stimulated to undergo activation characteristic of an immune response are, for example:

A) Antigen specific stimulation of immune responses. Either pre-immunized or naïve mouse splenocytes can be generated by standard procedures. In addition, antigen-specific T cell clones and hybridomas (*e.g.*, MBP-specific), and numerous B cell lymphoma cell lines (*e.g.*, CH27), have been previously characterized are available for the assays discussed below. Antigen specific splenocytes or B-cells can be mixed with specific T-cells in the presence of antigen to generate an immune response. This can be performed in the presence or absence of CLASP-2 to assay whether CLASP-2 modulates the immune response as measured by any of the assays in section 5.14.2.

B) Non-specific T cell activation. The following methods can be used to activate T cells in the absence of antigen: 1) cross-linking T cell receptor (TCR) by addition of antibodies against receptor activation molecules (*e.g.*, TCR, CD3, or CD2) together with antibodies against co-stimulator molecules, for example anti-CD28; 2) activating cell surface

receptors in a non-specific fashion using lectins such as concanavalin A (con A) and phytohemagglutinin (PHA); 3) mimicking cell surface receptor-mediated activation using pharmacological agents that activate protein kinase C (*e.g.*, phorbol esters) and increase cytoplasmic Ca^{2+} (*e.g.*, ionomycin).

C) Non-specific B cell activation: 1) application of antibodies against cell surface molecules such as IgM, CD20, or CD21. 2) Lipopolysaccharide (LPS), phorbol esters, calcium ionophores and ionomycin can also be used to by-pass receptor triggering.

D) Mixed lymphocyte reaction (MLR). Mix donor PBMC with recipient PBMC to activate lymphocytes by presentation of mismatched tissue antigens, which occurs in all cases except identical twins.

E) Generation of a specific T cell clone or line that recognizes a particular antigen. A standard approach is to generate tetanus toxin-specific T cells from a donor that has recently been boosted with tetanus toxin. Major histocompatability complex- (MHC-) matched antigen presenting cells and a source of tetanus toxin are used to maintain antigen specificity of the cell line or T cell clone (Lanzavecchia, A., *et al.*, 1983, Eur. J. Immun. 13: 733-738).

The anticipated mechanism of action of a CLASP-2 polypeptide or polynucleotide should define the appropriate assay to use to investigate its potential enhancement or inhibition of lymphocyte activation. For example, soluble proteins containing the CLASP extracellular domain may interfere with the interaction between T cells and antigen presenting cells. Such interaction plays a role in the MLR and in antigen-specific T cell activation, but not in non-specific T or B cell activation. The assays described above have the advantage of several possible detection methods for quantitation.

5.14.2. Methods for Generating an Immune Response in vivo

In various assays, an effect of an agent on immune cells is detected using an *in vivo* assay. The degree of an immune response can be measured or quantitated by a number of standard assays including those described below.

(A) Animal Model for Transplantation Rejection: Ectopic Heart Transplantation

In one embodiment, a standard animal model for graft versus host rejection is ectopic heart transplantation (Fulmer *et al.*, 1963, Am. J. Anat. 113: 273-281). This method involves using BALB/C mice (either sex, and range from 1–9 months) for transplanting cardiac tissue into a surgically-created pocket on the dorsum for both ears made by slitting

the skin over the auricular artery at the base of the ear. Small curved forceps are forced into the slit, bluntly dissecting between the skin and the cartilage plate. Donor tissue is eased into the base of the pocket near the distal edge of the ear. The auricular artery is used to seal off the opening of the pocket. Within 10 to 14 days pulsatile activity of the transplant should be observed. Gross appearance of the graft, patterns of vacuolar supply to the graft area and pulsatile activity can be easily observed utilizing transilluminated light during the first three weeks post-transplantation. Follow-up can continue for for several months.

(B) Animal model for Autoimmune Disease: Induction of Collagen Induced Arthritis (CIA)

Collagen Induced Arthritis (CIA) is a standard model for studying progression and immune (Courtenay *et al.*, 1980, Nature 283: 666 and Wooley *et al.*, 1981, J. Exp. Med. 154: 688). DBA/a mice can be used as an assay for the in vivo relevance of CLASP-2 in vitro testing potential immune therapeutics. In vivo experiments will be performed to examine the ability of potential therapeutics to prevent CIA. We will use 3-5 mice per group to statistically justify our results.

Once a titer of the potency of collagen type II (CII) is obtained therapeutics can be tested. In one embodiment, three mice will be immunized with three different concentrations of CII 50, 200, and 400 μg per animal (Nabozny *et al.*, 1996, J. Exp. Med., 183: 27-37). To induce CIA, animals can be immunized with an appropriate concentration of CII, determined as described above. One half of a 1:1 ratio of antigen:CFA can be injected at the base of the tail and the remainder equally divided in each hind footpad. Mice can be carefully monitored every day for the onset and progression of CIA thoughout the experiment until its termination 12 weeks post-immunization with CII. The pieces of heart transplanted can be approximately 3 X 3 mm in size. The severity of arthritis can be assessed following standard procedures known to one of skill in the art.

5.14.3 Assay Quantitation

(A) Tyrosine phosphorylation

Tyrosine phosphorylation of early response proteins such as HS1, PLC-r, ZAP-76, and Vav is an early biochemical event following T cell activation. The tyrosine phosphorylated proteins can be detected by Western blot using antibodies against phosphorylated tyrosine residues. Tyrosine phosphorylation of these early response proteins can be used as a standard assay for T cell activation (J. Biol. Chem., 1997, 272(23): 14562-14570). Any change in the phosphorylation pattern of these or related proteins when immune

responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(B) Intracellular Calcium Flux

The kinetics of intracellular Ca^{2+} concentrations are measured over time after stimulation of cells preloaded with a calcium sensitive dye. Upon binding the Ca^{2+} indicator dye, Fluor-4 (Molecular Probes), exhibits an increase in fluorescence level using flow cytometry, solution fluorometry, and confocal microscopy. Any change in the level or timing of calcium flux when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response

(C) Regulation of early activation markers

Increased and diminished expression/regulation of early lymphocyte activation marker levels such as CD69, IL-2R, MHC class II, B7, and TCR are commonly measured with fluorescently labeled antibodies using flow cytometry. All antibodies are commercially available. Any change in the expression levels of lymphocyte activation markers when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(D) Increased metabolic activity/acid release

Activation of most known signal transduction pathways trigger increases in acidic metabolites. This reproducible biological event is measured as the rate of acid release using a microphysiometer (Molecular Devices), can be used as an early activation marker when comparing the treatment of cells with potential biological therapeutics (McConnell, H.M. *et al.*, 1992, Science 257: 1906-1912 and McConnell, H.M., 1995, Proc. Natl. Acad. Sci. 92: 2750-2754). Any statistically significant increase or decrease in acid release of CLASP-2-treated sample, as compared to control sample (no treatment), suggest and effect of CLASP-2 on biological function.

(E) Cell proliferation/cell viability assays

(1) ^{3}H-thimidine incorporation

Exposure of lymphocytes to antigen or mitogen in vitro induces DNA synthesis and cellular proliferation. The measurement of mitotic activity by ^{3}H-thimidine incorporation into newly synthesized DNA is one of the most frequently used assays to quantitative T cell activation. Depending on the cell population and form of stimulation used to activate the T cells, mitotic activity can be measured within 24-72 hrs. in vitro, post ^{3}H-thimidine pulse (Mishell, B. B. and S. M. Shiigi, 1980, Selected Methods in Cellular Immunology, W. H. Freeman and Company and Dutton, R. W. and Pearce, J. D., 1962,

Nature 194: 93). Any statistically significant increase or decrease in CPM of CLASP-2-treated sample, as compared to control sample (no treatment), suggest and effect of CLASP-2 on biological function.

(2) MTS [5-(3-carboxymethoxyphenyl)-2-(4,5-dimethylthiazolyl)-3(4-sulfophenyl)tetrazolium, inner salt] is a colorimetric method for determining the number of viable cells in proliferation or cytotoxicity assays (Barltrop, J.A. *et al.*, 1991, Bioorg. & Med. Chem. Lett. 1: 611). 1-5 days after lymphocyte activation, MTS tetrazolium compound, Owen's reagent, is bioreduced by cells into a colored formazan product that is soluble in tissue culture media. Color intensity is read at 490 nm minus 650 nm using a microplate reader. Any statistically significant increase or decrease in color intensity of CLASP-2-treated sample, as compared to control sample (no treatment), can suggest an effect of CLASP-2 on biological function (Mosmann, T., 1983, J. Immunol. Methods 65: 55 and Barltrop, J.A. *et al.* (1991)).

(3) Bromodeoxyuridine (BrdU), a thymidine analogue, readily incorporates into cells undergoing DNA synthesis. BrdU-pulsed cells are labeled with an enzyme-conjugated anti-BrdU antibody (Gratzner, H.G., 1982, Science 218: 474-475.). A colorimetric, soluble substrate is used to visualize proliferating cells that have incorporated BrdU. Reaction is stopped with sulfuric acid and plate is read at 450 nm using a microplate reader. Any statistically significant increase or decrease in color intensity of CLASP-2-treated sample, as compared to control sample (no treatment), suggest an effect of CLASP-2 on biological function.

(F) Apoptosis by Annexin V

Programmed cell death or apoptosis is an early event in a cascade of catabolic reactions leading to cell death. A lose in the integrity of the cell membrane allows for the binding of fluorescently conjugated phosphatidylserine. Stained cells can be measured by fluorescence microscopy and flow cytometry (Vermes, I., 1995, J. Immunol. Methods. 180: 39-52). In one embodiment, any statistically significant increase or decrease in apoptotic cell number of CLASP-2-treated sample, as compared to control sample (no treatment), suggest an effect of CLASP-2 on biological function. For evaluating apoptosis in situ, assays for evaluating cell death in tissue samples can also be used in vivo studies.

(G) Quantitation of cytokine production

Cell supernatants harvested after cell stimulation for 16-48 hrs are stored at -80°C until assayed or directly tested for cytokine production. Multiple cytokine assays can be performed on each sample. IL-2, IL-3, IFN-γ and other cytokine ELISA Assays are

available for mouse, rat, and human (Endogen, Inc. and BioSource). Cytokine production is measured using a standard two-antibody sandwich ELISA protocol as described by the manufacturer. The presence of horseradish peroxidase is detected with 3, 3'5, 5' tertamethyl benziidine (TMB) substrate and the reaction is stopped with sulfuric acid. The absorbency at 450 nm is measured using a microplate reader. Any statistically significant increase or decrease in color intensity of CLASP-2-treated sample, as compared to control sample (no treatment), suggest an effect of CLASP-2 on biological function.

(H) NF-AT can be visualized by Immunostaining

T cell activation requires the import of nuclear factor of activated T cells (NFAT) to the nucleus. This translocation of NF-AT can be visualized by immunostaining with anti-NF-AT antibody (Cell 1998, 93: 851-861). Therefore, NF-AT nuclear translocation has been used to assay T cell activation. Similarly, NF-AT/luciferase reporter assays have been used as a standard measurement of T cell activation (MCB 1996, 12: 7151-7160).

(I) ELISA for collagen type II (CII)-specific antibodies (see above for related in vivo assay)

C(II) titers from serum of animals immunized with CLASP-2 can be measured and compared. Both TH1-dependent IgG2a and TH2-dependent IgG1 and IgE CII-specific antibody isotypes will be measured by ELISA. Mouse blood will be obtained by orbital bleed one and two months post-immunization with CII. Samples will be allowed to coagulate and centrifuge to obtain sera, and stored at –80°C until assayed by ELISA. Coat ELISA plates with CII and dilute sera. HRP conjugated goat, isotype specific antibody. Plates are then expose to TMB substrate and read at 450 nm using a microplate reader (Nabozny *et al.*, 1996, J. Exp. Med. 183: 27-37). Any change in the levels of Collagen specific antibodies by colorimetric test when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(J) Antibody Production by ELISPOT Assay

A solid-phase enzyme-linked immunospot (ELISPOT) assay for the quantification of isotype-specific antibody secreting cells (Czerkinsky *et al.*, 1983, J Immunol. Methods. 65: 109-121). Both human and mouse B cells can be tested for isotype and antigen specific antibody production. Although based on a standard ELISA, this technique becomes more sensitive by detecting antibody secretion from single cells. Any change in ELISPOT levels when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(K) Cellular degranulation following IgE cross-linking.

Two cell lines have been obtained from ATCC (MEG01 and HEL-17.92), both of which express the human FCεR1 receptor. FCεR1 is the high affinity receptor for IgE complexes, which when coupled to biotin can be cross-linked with avidin to induce degranulation and histamine release of lymphocytes. Following acylatation of the sample, histamine is quantified with an enzyme immunoassay competition assay (Immunotech). Histamine release. A statistically significant increase or decrease in histamine concentration of a CLASP-2 treated sample, as compared to control sample (no treatment), suggest an effect of CLASP-2 on biological function. Any change in frequency of degranulation or histamine levels when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(L) Cellular phenotyping of lymphocytes by flow cytometry and Immunocytochemistry

Determining the tissue distribution of lymphocytes following a pathological disorder can aid in identifying specific organ, tissue and lymphocyte involved in an immune response. Cellular phenotyping of lymphocyte trafficking is generally performed with by flow cytometry and Immunocytochemistry. There are several cluster determination (CD) molecules that are routinely used to identify phenotype, activation kinetics, and regulation events of cells. Any change in levels or distribution of CD molecules when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

(M) Structure/Function Assays: Homotypic and/or Heterotypic, Calcium-dependant Cell Adhesion

L929 cells can be transfected with CLASP-2 and Neomycin. G418-resistant clones can be screened for CLASP-expression with anti-CLASP peptide-specific antibodies. These CLASP-expressing clones can then be used to test for homotypic and/or heterotypic calcium dependent cell adhesion using the "cell aggregation assay" described for cadherin molecules (Murphy-Erdosh, C. *et al.*, 1995, J. Cell Biol. 129: 1379-1390). Any change in the levels of cellular aggregation when immune responses are generated in the presence of CLASP-2 is indicative of a CLASP-2 modulation of this response.

***

The following cDNA clones described in the Specification and further described in the Examples below have been deposited with the American Type Culture

Collection, 10801 University Boulevard, Manassas, VA 20110-2209 under the Budapest Treaty on March 24, 2000 and given the Accession Nos. indicated:

hCLASP-2A 3' clone (AVC-PD1) ATCC accession number PTA-1563

hCLASP-2A 5' clone (AVC-PD2) ATCC accession number PTA-1562

hCLASP-2B clone (AVC-PD12) ATCC accession number PTA-1573

The following cDNA clones described in the Specification and further described in the Examples below have been deposited with the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209 under the Budapest Treaty on _____ and given the Accession Nos. indicated:

hCLASP-2 clone hC2GR3.3 (AVC-PD14) ATCC Accession No._____

hCLASP-2 clone hC2RT (AVC-PD19) ATCC Accession No._____

***

6. EXAMPLES

EXAMPLE 1

Cloning of CLASP-2

The cloning of the CLASP gene family has not been a straightforward process. The cloning of each CLASP family member required the use of multiple techniques and resources. CLASP-2 was cloned in the following manner: an expressed sequence tag or EST clone (IMAGE clone 815795, derived from human germinal B cells) was identified based on a BLAST search of human GenBank human EST database using CLASP-1 sequences. IMAGE clone 815795 was sequenced completely. A polynucleotide probe prepared from 815975 sequence was labeled with ^{32}P-dCTP and used to screen human cDNA libraries including Jurkat (Stratagene) and Ramos B cell cDNA library (James Boulter, UCLA). The screening methods employed were as described in Maniatis *et al.*, 1989, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Several clones were identified and clone C9, with an insert of 3,752 base pairs, was sequenced (ABI dye-sequencing system, PE Applied Biosystems; Perkin-Elmer Corporation, 761 Main Avenue, Norwalk, CT, U.S.A.). A 5' probe was prepared from C9 sequence and used to rescreen the cDNA libraries. Several clones were isolated, but could not be excised from the phage (Stratagene, CA) without deleting the insert. To circumvent this problem, anchor PCR was performed using M13F primer and CLASP-2 primer (C96AS). The PCR fragment was cloned using the pGEM-T system (Promega), although initial attempts were unsuccessful. The isolated sequence encompassed additional but incomplete cDNA sequence and was determined to carry at least one mutation that may have allowed it to be propagated in bacteria. Commercial libraries from multiple tissue sources including human placenta, B cell, T cell and peripheral blood were exhaustively screened and re-screened resulting in the acquisition of only partial cDNAs. Generation of cDNA libraries using oligo dT or CLASP-specific primers also resulted in the acquisition of partial cDNAs. Genomic libraries were screened to obtain a portion of the genomic locus for each of the CLASP genes, and a genomic walk was initiated to obtain 5' exons and extend the cDNA sequence.

To obtain additional 5' CLASP-2 sequence, portions of the cDNA and genomic sequence from a BAC (Bacterial Artificial Chromosome) genomic library were compared to the NCBI database by BLAST. A genomic clone (Genbank identifier: gi9988160) comprising random, shotgun genomic sequence was identified. Using TFASTX (Pearson and Lipman, PNAS (1988) 85:2444-2448), the amino-terminal sequence of human

CLASP4 was compared to 6 frame translation of gi9988160. Areas of gi9988160 that encoded amino acids with high similarity to CLASP4 amino acid sequence were used to design CLASP-2-specific oligonucleotides for RTPCR (reverse transcriptase polymerase chain reaction according to manufacturers instructions: Reverse transcriptase Gibco/BRL, Taq Polymerase from Sigma). Using oligonucleotides hC2gS5 (nucleotides -66 to –44 of FIG. 11) and C2AS18 (reverse complement of nucleotides 2120 to 2140 of FIG. 11) an RTPCR product of approximately 2.2kb was generated, sequenced (dideoxynucleotide termination sequencing, Beckman Coulter CEQ2000) and shown to be additional human CLASP-2 5' sequence. Further complicating the cloning full-length CLASP cDNA products was the difficulty to clone (and subclone) certain CLASP cDNA products. Standard isolation of some of the CLASP cDNAs from a pure phage population following screening of commercially available cDNA libraries ("ZAP-out" procedure, Stratagene) resulted in no bacterial colonies. Similarly, certain RT-PCR products could not be cloned into standard plasmid vectors. No colonies were isolated by cloning these fragments into vectors lacking promoters, reverse orientations, low copy vectors, or by growth at altered temperatures or levels of antibiotic for plasmid selection (examples: CLASP-7 - HC7gS6 to HC7gAS1 and HC7gS3 to HC7AS14; CLASP-4 – C4P2 to hC4ASTM and C4P2 to HC4AS3'; CLASP-1 – hC1S5' to hC1AS3'Kpn and C1S7 to hC1AS3'Kpn; see Primer Table below). One possibility is that sequences contained within certain regions of CLASP cDNAs are bacteriacidal and therefore not amenable to cloning. To circumvent these problems direct sequencing of RT-PCR products was performed.

Primer Table

CLASP gene	Sense Primer	Sense sequence	Antisense Primer	Antisense sequence
CLASP-7	HC7gS5	AGGCCTTGTCTCTGTTTACCTG	HC7gAS1	TGTCATGTACTGCACTCGCACAGC
CLASP-7	HC7gS3	ACAGGAACCTGCTGTACGTGTAC	HC7AS14	TCGTGGCTGCACAGGATGCGGGTG
CLASP-4	C4P2	GACCCATTAGGAGGTCTAC	HC4AS3'	CGGGATCCATTGTCACCGTACATCT GC
CLASP-4	C4P2	GACCCATTAGGAGGTCTAC	HC4AS3'	CGGGATCCATTGTCACCGTACATCT GC
CLASP-1	hC1S5'	TATGTCTCAGTCACCTACCTG	HC1AS3'Kpn	CTTGGTACCACTTCAGCACTAGATG AGATG
CLASP-1	C1S7	TCAAGACCAGGGCATGCAAG	HC1AS3'Kpn	CTTGGTACCACTTCAGCACTAGATG AGATG

In-frame stop codons were not present suggesting that the cDNA was not full length. To obtain the 5' terminus of CLASP-2 5' RACE was employed. Antisense oligonucleotides directed against the 5' end of the longest CLASP-2 sequence were generated:

Primers used for human CLASP-2 5' RACE

primer	sequence(5' TO 3')	nucleotide position
HC2RACE1		
	AAGAGCAGCATCTCCCGTAAACAGTC	-15 to 11
HC2RACE2		
	TAACAAGCTCTGTGCTTCCTCTTCCG	414 to 443
HC2RACE3		
	ACCACTTTGTTCGGAAGCTGTCGAAACTC	512 to 540
HC2RACE4		
	TTTGTACAGCCAGCCATGCTTGGTGATC	634 to 661

RACE was carried out using Generacer kit (Invitrogen) according to manufacturers specifications using polyA selected mRNA from 9D10 B cell tissue culture line. The sequence of the oligonucleotides presented is the reverse complement (*i.e.*,

antisense) of the the CLASP1 cDNA at the indicated position based upon numbering in FIG. 11.

The full length cDNA (presented in FIG. 11) is therefore a compilation of cDNA from cDNA libraries, RTPCR products and 5' RACE products. The sequence of the CLASP-2 cDNA is shown in FIG. 11.

EXAMPLE 2

Tissue and Cell Line Expression of the CLASP-2 gene

Multiple Tissue Northern blots were purchased from Clontech; hybridization procedures were followed according to manufacturer's procedures and recommendations. Human T cell line (Jurkat), human myelomonocyte cells (MV4-11), B cells (9D10), monocytes (THP-1), mouse T cells (3A9), mouse B cells (CH27), human promyelocyte (HL60) and human kidney epithelial cells (293 cell line) were maintained as cultured cell lines. For Multiple Cell Northerns, RNA was prepared from cell suspensions using the GIBCO-BRL Trizol system. All steps were performed according to the manufacturer's procedures and recommendations. RNA concentrations were determined by the 260nm/280nm light absorption of the RNA solution. 20 μg RNA was ethanol precipitated and resuspended in formamide/formaldehyde buffer and incubated for 15' at 65°C to eliminate putative secondary structures. RNA samples were run over night on a 1.1% agarose gel containing 1.5% formaldehyde (both gel and running buffer were 20 mM sodium phosphate, pH 7.5). To visualize RNA after gel migration, approx. 0.5 μg ethidium bromide was added to each sample prior to the run together with RNA loading buffer. RNA in the gel was then visualized by 260nm wavelength light. After soaking the gel for 15' in deionized water to reduce the concentration of ethidium bromide in the gel, the RNA was transferred onto Amersham Hybond-N plus membrane by capillary blotting in 20 x SSC buffer for 5 hours. Subsequent to blotting, the membrane was washed in 5 x SSC for 3' and RNA was crosslinked to the membrane by UV light (Stratagene Stratalinker).

A probe which recognizes CLASP-2 isoforms A, B, C, and D (probe HC2.2) was used. Probe HC2.2 encompasses to nucleotides 3920 to 4650 (731 bp long) of CLASP-2A. The HC2.2 probe was prepared using standard labeling kits and desalted using pasteur pipette G-50 Sephadex column in TEN (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 100 mM NaCl).

Hybridizations of ^{32}P dCTP labeled DNA probes to the membrane bound RNAs (multiple tissue and multiple cells) were carried out in CLONTECH EXPRESSHYB

solution, at 68°C and for 1-2 hours. Blots were washed 2 times in 2x SSC 0.1% SDS for 10' each at 50°C and then twice in 0.2 x SSC 0.1% SDS for 10' each at 50°C, followed by a 5' wash in 2xSSC at 50°C. Exposure to KODAK BIOMAX MS film was carried out at minus 80°C using amplifying screens. Typical exposure times were 10 to 36 hours.

EXAMPLE 3

Southern Analysis of CLASP-2

BAC DNA was prepared from E. coli over night cultures using the QIAGEN DNA preparation system. All preps were performed according to the manufacturer's procedures, including the modifications for low copy number DNA constructs. Genomic DNA was prepared from HeLa cells (ATCC #CCL-17) using the methods described by Sambrook, Fritsch and Maniatis (1989); DNA concentrations were determined by the 260nm light absorption of the DNA solution, and aliquots corresponding to 20 microgram (μg) genomic DNA or 2 μg for BAC DNA were used for restriction enzyme digests with Eco RI or HinD III (genomic DNA) or Eco RI and Pst I (BAC DNA). Digests were carried out in 150 microliter volume for 4 hours at 37°C. Digested DNA was ethanol precipitated and the pellet was resuspended in 20 microliter deionized water prior to migration over a 1.2 % agarose gel at 35 V over night. Running buffer was TAE, and the gel contained 0.1 μg ethidium bromide/ml to visualize DNA.

Subsequent to gel separation, DNA was visualized by 260 nm wavelength light. The gel was then washed twice for 20' in denaturing buffer (0.5M NaCl, 0.4 N NaOH) and twice in neutralization buffer (1.5 M NaCl, 0.5 M TRIS pH 8.0). DNA was transferred from the gel onto AMERSHAM HYBOND N membrane by capillary blotting in 20 x SSC for 5 hours. The DNA was crosslinked to the membrane by UV light using a Stratagene Stratalinker.

A probe, HC2.1, which recognizes CLASP-2, was used. Probe HC2.1 encompasses nucleotides 325 to 1126 (802 bp long) of CLASP-2A. The HC2.1 probe was prepared using standard labeling kits and desalted using pasteur pipette G-50 Sephadex column in TEN (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 100 mM Nacl). Hybridizations of ^{32}P dCTP labeled DNA against DNA immobilized onto the membrane were carried out at 65°C overnight in modified CHURCH hybridization solution (7% SDS, 0.5 M sodiumphosphate, 1mM EDTA). Membranes were then exposed to KODAK BIOMAX MS film at minus 80°C. Typical exposure times were 12 hours for genomic DNA southern analysis and 3 hours for BAC DNA Southern analysis.

The genomic DNA southern analysis revealed two fragments (~4.5 kb and 1.85 kb) in the Eco RI digested DNA but three fragments in BACs 4 and 6 DNA. The two major bands are identical in both genomic and BAC DNA (FIG. 7).

EXAMPLE 4

CLASP-2 Genomic Cloning

Genomic clones of human CLASP-2 were obtained using the Release I high density filters from Genome Systems Inc (cat # FBAC-4434). Two rounds of screening were completed. The first round of screening was carried out using a probe corresponding to nucleotides 3830 to 4558 of the human CLASP-2 cDNA by standard protocols specific by Genome Systems. This screen identified two genomic clones, referred to as AVC BAC4 and 7. A second round of screening using a probe that corresponded to nucleotides 1208 to 1604 of human CLASP-2 cDNA identified clone AVC BAC26. All the clones were partially sequenced to authenticate that they were indeed CLASP-2 genomic clones, to verify exon sequences, and to identify exon/intron boundaries. Oligonucleotides for sequencing the BACs were based upon human CLASP-2 cDNA sequence. Sense and antisense sequencing oligonucleotides were designed along the length of the human CLASP-2 cDNA spaced approximately every 200 nucleotides to ensure a high density of coverage of the corresponding genomic regions. Sequencing reactions with primers and BAC DNA were carried out by standard PCR sequencing using Big Dye termination sequencing mix (ABI). Results from sequence reactions were analyzed using Sequencher software (Genecodes). The results are summarized in FIG. 6.

EXAMPLE 5

Expression of Recombinant CLASP-2A Polypeptide in Bacterial Cells

Portions of hCLASP-2 were cloned into the GST expression vector pGEX (Pharmacia). These include the region spanning the potential Cadherin processing site through 200 amino acids of the predicted extracellular domain (nucleotide 866 – 1459; GST-EC12; 55 kD fusion) and a portion of the intracellular domain (nucleotide 3230 – 4065; GST-cyto; 57 kD fusion). These regions were amplified using primers at the limits of these sequences on either cDNA clones or cDNA generated from Jurkat or Human Peripheral Blood RNA. Amplified DNA sequences were digested with restriction enzymes for cloning in-frame into GST expression vectors. Fusion proteins were expressed by IPTG induction in DH5α and purified according to instructions from Pharmacia using glutathione-Sepharose

(Pharmacia). SDS-PAGE gel stained with Coomassie Blue showing induced and uninduced expression of the GST-CLASP-2-cyto construct is shown in FIG. 8. These recombinant proteins were expressed in DH5α and purified according to instructions from Pharmacia using glutathione-Sepharose. Such recombinant proteins were used to generate antibodies (Josman laboratories) using a AVC Rapid Immunization Protocol.

The full length CLASP can easily be expressed from either the beginning of the hCLASP-2 sequence (in frame with nucleotide 2) or from the first or second methionine (nucleotide 278 or nucleotide 476, underlined in FIG. 1) through to the stop codon (nucleotide 4058). Assuming that the GST moiety has a weight of 26 kD, the total predicted sizes are 180, 168, and 164.5 kD respectively. Alternatively, other bacterial expression systems such as 6CLASP HIS tags, Calmodulin binding protein, maltose binding protein can also be used in a similar manner.

EXAMPLE 6

Expression of Recombinant CLASP-2A Polypeptide in Mammalian Cells

Example 6A. Secreted fusions

Several portions of the predicted extracellular domain were constructed as hIgG fusions using the CD5gamma-1 expression vector (kindly provided by B. Seed, Harvard University). Polypeptides were cloned into this vector in frame with a CD5 leader sequence that directs the fusion protein into the secretory pathway and in frame with a C-terminal hIgG(Fc) protein. This fusion can be secreted from cell lines such as 293 (Hsieh, J-C., 1999, Nature 398: 431-436). Sense primers with hCLASP-2 sequences beginning at nucleotide 866 and antisense primers at nucleotide 1459 (EC12-IgG), nucleotide 2389 (ECC-IgG) and nucleotide 2857 (ECM-IgG) were used to amplify portions of the extracellular domain for insertion into this vector. Recombinant vectors were purified by Maxiprep (Qiagen) and transfected into 293 EBNA- T cells (kindly provided by B. Seed, Harvard University) by calcium phosphate techniques (Sambrook and Maniatis). After 2-7 days, secreted expression was analyzed by an ELISA against the hIgG fusion using a goat F(ab')2 anti human IgG(Fc) antibody (Jackson Immunolabs) and Protein-A-HRP (Pierce). Intracellular expression was monitored by immunofluorescence microscopy with a FITC labeled goat anti Human IgG(Fc) antibody (Caltag).

Example 6B. Intracellular fusions

Similar methods have been used to construct fusions for expression of full length hCLASP-2 isoforms as well as truncated C-terminal forms in other cell lines such as Jurkat. Recombinant hCLASP-2 fragments were either isolated by digestion of cDNA clones or amplified by primers flanking specific regions (Please provide some specific regions). These can be cloned into expression vectors such as pBJ1-neo (Mark Davis, Stanford University), Peak12 (B. Seed, Harvard University), and pDsRed1-N1 (Clontech). pBJ1-neo and Peak12 allow untagged expression of recombinant proteins and pDsRed1-N1 will allow either untagged or a C-terminal Red fluorescent protein tag. These can be used to generate protein or for expression of various forms for functional analyses.

EXAMPLE 7

Antisense Inhibition of CLASP-2 Expression

Example 7A. Inhibition of CLASP-2 expression *in vitro*

In this example, inhibition of CLASP-2 expression is examined using an *in vitro* cell-free expression system. To identify the useful antisense oligonucleotides, a series of antisense phosphorothioate oligonucleotides (PS-ODNs), which span portion CLASP-2 sequence, can be systematically assayed for the ability to block CLASP-2 expression *in vitro*.

For inhibition of CLASP-2 expression *in vitro*, a CLASP-2 transcription/expression plasmid can be used according to standard methodology for *in vitro* transcription and translation of sense CLASP-2 RNA. Coupled transcription-translation reactions can be performed with a reticulocyte lysate system (Promega TNTTM) according to standard conditions. Each coupled transcription/translation reaction can include CLASP-2 RNA transcribed from the expression plasmid, and a test antisense polynucleotide at a range of standard test concentrations, as well as the luciferase transcription/translation internal control to normalize each reaction (see, *e.g.*, Sambrook *et al.*, *supra*, Ausubel *et al.*, *supra*). The translation reaction can also be performed with sense CLASP-2 RNA that is synthesized *in vitro* in a separate reaction and then added to the translation reaction. ^{35}S-Met is included in the reaction to label the translation products. The negative control is performed without added PS-ODN or a sense PS-ODN.

The labeled translation products can be separated by gel electrophoresis and quantitated after exposing the gel to a phosphorimager screen. The amount of CLASP-2 protein expressed in the presence of CLASP-2 specific PS-ODNs can be normalized to the co-expressed luciferase control.

Example 7B. Inhibition of CLASP-2 expression *ex vivo*

A. Reagents

Cells: Jurkat, Clone E6-1 ATCC TIB-152; 9D10 ATCC CRL8752; additional cells from the ATCC or NCI.

Media and solutions: RPMI 1640 medium, BioWhitaker; DMEM/M199 medium, BioWhitaker; EMEM, BioWhitaker; Fetal Bovine Serum, Summit (stored frozen at -20°C, stored thawed at 4°C); Trypsin-EDTA, GIBCO (catalogue #25300-054) (stored frozen at -20°C, stored thawed 4°C; Isoton II (stored at RT); DMSO (stored at RT); oligonucleotides (see Table 1 and FIG. 3, stored in solution at -20°C); PBS (Ca^{2+}/Mg^{2+} free); TE; 10 mM Tris-HCL, pH 8.0; 1 mM EDTA.

To prepare oligonucleotide stocks: Oligonucleotide nucleotides (PS-ODNs) can dissolved in the appropriate amount of TE to make a concentrated stock solution (1 - 20 mM).

B. Treatment of cells ex vivo with antisense CLASP-2 oligonucleotides

Stock cultures of cells in log-phase growth (in T75 flask) can be used. Jurkat, and 9D10 cells are used in this assay. Jurkat and 9D10 are suspension cultures and are passed through dilutions in media. Cell density is measured using a Coulter counter or hemacytometer.

For 6-well dishes, 1.1 x 10^5 cells total per well, 2 ml/well is added. The amount of cells can be scaled up or down proportionally for 12-well, 100 mm, or 150 mm dishes. For example, for 12-well dishes, use 4.6 x 10^4 cells in 2 ml media; for 100 mm dishes use 6 x 10^5 cells in 10 ml media; for 150 mm dishes use 1.7 x 10^6 cells in 35 ml media.

An appropriate number of cells (as described in step 2 above) are collected, centrifuged and resuspended in media containing a range of ODN concentrations. The cells are treated in single, duplicate, or triplicate wells. Control wells are treated with TE or sense ODNs diluted in media.

The suspension cultures are washed and resuspended daily with PS-ODN media.

Suspension cultures are grown for 2-4 days. Cells are washed with PBS and density measured using a Coulter counter or a hemocytometer. If necessary, the cells are replated at 1.1 x 10^5 cells per well, 2 ml media per well, and fed with PS-ODN as described above.

Samples of the cells can also be harvested for analysis to determine the effects of CLASP-2 antisense ODNs. Samples are harvested for RNA and analyzed by either Northern analysis or RT-PCR for the presence of CLASP-2 mRNA. Functional consequences of CLASP-2 antisense ODNs can be analyzed by measuring the ability of Jurkat and 9D10 cells to be activated. Jurkat cells are activated by exposure to anti-CD3 and anti-CD28 crosslinking antibodies, and 9D10 cells are activated by exposure to anti-IgM crosslinking antibody or P. aeruginosa lipopolysaccharide. A hallmark of activation, calcium influx, can be measured by flow cytometry. Additionally, ELISA assays can be used to measure Interleukin-2 production from Jurkat cells and secreted IgM can be measured using standard assays from 9D10.

Table 5 below shows exemplary oligonucleotides for this assay:

Table 5

Oligo	Sequence 5'- 3'	length	notes/comments
1	GAAGGCGATCATCACGT GGCCTTCCATCGC	30-mer	encompasses nucleotides 473-502 and spans the putative initiator methionine (underlined). The function of HC2A, 2B, 2C, and 2E isoforms can be eliminated by this oligonucleotide.
2	GCTTCAAGTAATGACTGG TGCAGAACATCTG	31-mer	Oligonucleotide that should recognize HC2A, 2B, 2D, 2E, and 2F. Encompasses nucleotides 2121-2151. Can be eliminate function of these CLASP-2 isoforms.
3	GCTCCTCCTCAGGCAGGC GCTATGGCTGTGG	34-mer	oligonucleotide specific for HC2C based upon a specific exon found at nucleotide 2927. Can eliminate only HC2D function.
4	GTAGGCCCGGTGCAGCGT GTCATACAGATGG	31-mer	oligonucleotide specific for HC2B, 2C, 2D and 2E based upon specific exon sequence found at nucleotide 3153. Can eliminate function of these CLASP-2 isoforms.
5	GCAATGTCTGAGACTTTC GATCATGAACTATG	32-mer	oligonucleotide specific for HC2A, 2B, 2E, and 2F. Encompasses nucleotides 1987-2018. Can eliminate function of these CLASP-2 isoforms.
6	CAGGAGCTGGTTCTTAAA	18-mer	oligonucleotide specific for HC2A, 2D and 2E. Encompasses nucleotides 2219-2224. Can eliminate function of these CLASP-2 isoforms

Table 5 legend. All nucleotide numeration are relative to Human CLASP-2A (HC2A). See FIG. 2A.

EXAMPLE 8

Example 8A. Synthesis of carboxyl-termini PDZ-ligand peptides

The GST-PDZ fusion proteins are made following standard procedures. An exemplary GST-PDZ fusion protein was constructed as follows: A 572 bp fragment encoding

two PDZ domains of the human neDLG gene (Genbank Accession No. U49089.1) was amplified from total Jurkat RNA by RT-PCR according to standard protocols (Sambrook, Fritsch, and Maniatis, 1989, Molecular Cloning – A Laboratory Manual. Cold Spring Harbor Press.) using primers flanked by restriction endonuclease sites for cloning. Fragments were purified by Sephaglas (Pharmacia), digested with the appropriate enzymes, and ligated into the GST expression vector pGEX-3X (Pharmacia) cut with similar enzymes. Recombinant constructs were confirmed by sequencing. Fusion proteins were expressed by IPTG induction in DH5α and purified using glutathione-Sepharose (Pharmacia) according to instructions from Pharmacia. Excess glutathione was removed using a PD10 desalting column (Pharmacia) and samples were diaconcentrated by placing the protein in dialysis tubing (14,000 MW cutoff) and laying the tubing on polyethylene glycol (3350; Sigma) until volume had been reduced by approximately 50%. Glycerol was then added to 35% final concentration and samples were stored at –20°C. These recombinant proteins have been used to generate antibodies (Josman laboratories) by standard protocols and for biochemical studies describe herein.

Synthetic peptides corresponding to the carboxyl-terminus of a protein of interest are synthesized by standard resin-based chemistry (*e.g.*, FMOC), labeled with biotin at the amino-terminus when indicated, and cleaved from the resin using a halide containing acid (*e.g.*, trifluoroacetic acid). The synthetic peptides are then purified by reverse phase high performance liquid chromatography (HPLC) and the identity of the peptides are confirmed by mass spectrometry.

Example 8B. Measurement of CLASP-2 peptide binding to PDZ Domain-containing proteins

The binding of a biotinylated carboxyl-terminal peptide to a GST-PDZ fusion protein is measured as follows:

(1) GST fusion protein containing one or more PDZ domain(s) is coated onto a protein-binding surface. The protein-binding surface is the surface of a polystyrene plate, which in some cases has been pre-treated by coating with 5 μg/ml of goat-anti-GST polyclonal antibody followed by blocking with excess bovine serum albumin (BSA). The concentration of GST fusion protein used is 5–10 μg/ml and the reaction of the GST fusion protein with the plate is carried out in PBS for 1 – 16 hours at

4°C. If not already blocked, the plate is then blocked with BSA (2% in PBS, 2 hours, 4°C)

(2) The plate is washed with PBS.

(3) The biotinylated peptide (generally 0.2–20 μM) is then added to the plate and allowed to react in PBS/2% BSA buffer with the GST fusion protein for 10 minutes at 4°C followed by 20 minutes at 25°C. In cases where competition between a labeled (biotinylated) and unlabeled (non-biotinylated) peptide is performed, the unlabeled peptide is added immediately prior to adding the labeled peptide. .

(4) The plate is washed with PBS.

(1) 0.5 μg/ml steptavidin-HRP conjugate is added to the plate in PBS/2 % BSA buffer and allowed to react for 20 minutes at 4°C.

(6) The plate is washed 5 X with detergent (tween 20) containing solution.

(7) The plate is developed by addition of HRP-substrate solution for 20 minutes at room temperature.

(8) The reaction of the HRP and its substrate is terminated by addition of 1 M sulfuric acid.

(9) The optical density of each well of the plate is read at 450 nm.

In cases where measurement of the apparent affinity of PDZ-ligand interaction is desired, the above procedure is carried out with multiple concentrations of the labeled peptide being used in a single experiment. A plot of binding versus peptide concentration added is then fit to the equation:

$$\text{Binding [peptide]} = \text{Saturation Binding} \times (\text{[peptide]} / (\text{[peptide]} + Kd))$$

where "Binding [peptide]" is the binding of a given concentration of peptide to the GST-PDZ fusion protein minus binding to the GST alone control, "Kd" is the apparent affinity of the binding reaction, and "Saturation Binding" is computed to allow the best fit of the data to the above equation. The term apparent affinity is used because the reaction may

not reach equilibrium during the duration of the binding reaction in which case the apparent affinity would underestimate the actual affinity (*i.e.*, actual Kd < observed Kd).

EXAMPLE 9

Expression of human CLASP -2 in activated T-cells

General experimental design

The expression profiles of human CLASP-2 in T cells upon T cell activation was determined by Northern analysis. Jurkat E6 lymphoblasts were activated by treatment with anti-CD28, PMA, and Ionomycin. Subsequently, total RNA was extracted from cell aliquots harvested at 0, 1, 2, 4, 8, and 14 hours post activation. The RNA concentration of each preparation was determined by the absorption at 260 nm using a spectrophotometer and concentrations of the different RNA preparations were adjusted such that equal quantities of each RNA preparation could be subjected to Northern analysis. Even gel loading was monitored by ethidium bromide staining of the formaldehyde-agarose gel. Northern membranes were hybridized to radioactively labeled probes corresponding to portions of human CLASP-2 and human beta-actin. Expression levels of CLASP-2 at different time points post T-cell activation are proportional to the radioactive signal generated by hybridization by the CLASP-2 specific radioactively labeled probe that remained bound to the Northern membrane under stringent washing conditions. The entire experiment was done in duplicate.

Jurkat E6 cell activation

Jurkat E6 cells were maintained and tested in complete IMDM medium supplemented with 2 mM glutamine, 10 mM HEPES, 100 u/mL penicillin, 100 µg/mL streptomycin, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate (Gibco/BRL), 50 µM beta mercaptoethanol (Sigma), and 10% fetal calf serum (Gemini). T cells were activated as described per Fraser et al., using 0.1 g/mL mouse anti-human CD28 monoclonal antibody (PharMingen International catalog number 33741A), 50 ng/mL PMA (Sigma), and 1 µM ionomycin (Calbiochem). Following incubation at 37°C and 5.0% v/v CO_2, 0.5 x10^6 cells were harvested by centrifugation at 500 x g for 10 minutes (min) at room temperature at 0, 1, 2, 4, 8 and 14 hours post activation and subjected to RNA extraction.

For RNA preparation, probe labelling and Northern analysis protocols, see methods and procedures described in Example 2 above. The CLASP-2 specific probe

encompassing nucleotides 5352 to 5922 was generated by PCR from a plasmid containing cloned CLASP-2 cDNA sequences using primers C2S12 and C2AS21.

Hybridization, Washing, and Exposure

Blots were washed twice in 2x SSC 0.1% SDS for 10 min each at 60° C and then twice in 0.2x SSC 0.1% SDS for 10 min each at 60° C, followed by a 5' wash in 2xSSC at 60° C. Exposure to KODAK BIOMAX MS film was carried out at minus 80° C using amplifying screens. Typically, exposure times were 10 to 36 hours. Signal intensities on Northern membranes were quantified by the use of a phosphor imager system (STORM, Molecular Dynamics). Signals were counted in the "volume report" mode.

Results

CLASP-2 expression levels as determined by Northern analysis (FIG. 14) slightly decrease at 1 hour post activation. The maximum decrease of approximately 36 % is seen at 2 hours post activation. Expression levels augment again at 4 hours post activation but do not attain the level that is seen before activation (0 hours). Intensities of CLASP-2-specific signals on the Northern blot were quantified by phosphor imager analysis. Rectangles were drawn around the areas of CLASP-2-specific signal and total quantity of signal was determined by the "volume report" mode; phosphor imager quantification results of two entirely independent experiments are shown in the diagram (green bars corresponds to Northern blot shown). The above result suggests, that transcriptional control of CLASP-2 expression and T-cell activation are functionally linked to each other.

EXAMPLE 10

Chromosomal location of CLASP-2 and possible disease associations

CLASP-2 cDNA sequences have been mapped to the genomic clone (GI:9926440, GI:9988160) by use of sequence homology bioinformatics tools BLAST.

Clone (GI:9926440, GI:9988160) has previously been mapped to the chromosomal location 13q12-q13. The literature research reports that the mutations, deletions, rearrangements, disomies and/or breakpoints (in general: chromosomal aberations) in below listed genes make the genes strong candidates for the onset of the listed diseases/disorders. Because the CLASP-2 gene is localized in the chromosome location 13q12-q13, abnormal CLASP-2 gene regulation or deletion, rearrangement and/or mutations in CLASP-2 locus might be directly or indirectly associated with the onset of the listed diseases. Further, CLASP-2 gene can be used as a genetic probe to detect the abnormality in

regions of these below listed genes and as a diagnostic marker for the related disease/disorders.

CANDIDATE GENES	*LOCUS*	*RELATED DISEASE/DISORDERS*
IPF1:Insulin promoter factor1	13q12.1	MODY4: non insulin-dependent juvenile type, Defect in pancreatic islet development and insulin transcription.
BRCA2	13q12.3	BCLL2: B cell lymphoma, deletion encompassing BRCA2 causes B cell lymphoma. BRCA2 is one of the responsible genes for DNA repairing in S phase.
	13q13.1-q14.3	Deletion of these locus causes MDS6: Myelo dysplastic syndrome type 6 including AML.

***

The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and any clones, DNA or amino acid sequences which are functionally equivalent are within the scope of the invention. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims. It is also to be understood that all base pair sizes given for nucleotides are approximate and are used for purposes of description.

All publications and patent documents cited above are hereby incorporated by reference in their entirety for all purposes to the same extent as if each were so individually denoted.

***

WHAT IS CLAIMED IS:

1. An isolated CLASP-2 polynucleotide, wherein said polynucleotide is (a) a polynucleotide that has the sequence of SEQ ID NO: 1, 3, 5 or 9; or (b) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and encodes a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10 or an allelic variant or homologue of a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10; or (c) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and encodes a polypeptide with at 25 contiguous residues of the polypeptide of SEQ ID NO: 2, 4, 6 or 10; or (d) a polynucleotide that hybridizes under stringent hybridization conditions to (a) and has at least 12 contiguous bases identical to or exactly complementary to SEQ ID NO: 1, 3, 5 or 9.

2. The polynucleotide of claim 1, wherein said polypeptide specifically binds to a PDZ domain of PSD95, DLG1 or neDLG.

3. The polynucleotide of claim 2, wherein said polypeptide has a binding affinity of at least 10^4 M^{-1} for binding PSD95, DLG1 or neDLG.

4. The polynucleotide of claim 1that encodes a polypeptide having the full-length sequence of SEQ ID NO: 2, 4, 6 or 10.

5. The isolated polynucleotide of claim 1, comprising the cDNA coding sequence of ATCC Deposit Nos. PTA-1562 and PTA-1563 and PTA-1573.

6. An isolated CLASP-2 polynucleotide comprising a nucleotide sequence that has at least 90% percent identity to SEQ ID NO: 1, 3, 5 or 9.

7. An isolated polypeptide comprising a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 2, 4, 6 or 10 and is immunologically crossreactive with SEQ ID NO: 2, 4, 6 or 10 or shares a biological function with native CLASP-2.

8. A vector comprising the polynucleotide of claim 1.

9. An expression vector comprising the polynucleotide of claim 1 in which the nucleotide sequence of the polynucleotide is operatively linked with a regulatory sequence that controls expression of the polynucleotide in a host cell.

10. A host cell comprising the polynucleotide of claim 1, or progeny of the cell.

11. A host cell comprising the polynucleotide of claim 1, wherein the nucleotide sequence of the polynucleotide is operatively linked with a regulatory sequence that controls expression of the polynucleotide in a host cell, or progeny of the cell.

12. The host cell of claim 10 which is a eukaryote.

13. The polynucleotide of claim 1 that is an antisense polynucleotide less than about 200 bases in length.

14. An antisense oligonucleotide complementary to a messenger RNA comprising SEQ ID NO: 1, 3, 5 or 9 and encoding CLASP-2, wherein the oligonucleotide inhibits the expression of CLASP-2.

15. An isolated DNA that encodes a CLASP-2 protein as shown in SEQ ID NO: 2, 4, 6 or 10.

16. The polynucleotide of claim 1 that is RNA.

17. A method for producing a polypeptide comprising:

(a) culturing the host cell of claim 10 under conditions such that the polypeptide is expressed; and

(b) recovering the polypeptide from the cultured host cell or its cultured medium.

18. An isolated polypeptide encoded by a polynucleotide of claim 1 (a) or (b).

19. The polypeptide of claim 18 that has the amino acid sequence of SEQ ID NO: 2, 4, 6 or 10, or a fragment thereof.

20. The isolated polypeptide of claim 18, wherein the polypeptide is cell-membrane associated.

21. The isolated polypeptide of claim 18, wherein the polypeptide is soluble.

22. The polypeptide of claim 19, wherein the polypeptide is fused with a heterologous polypeptide.

23. An isolated CLASP-2 protein having the sequence as shown in SEQ ID NO: 2, 4, 6 or 10.

24. A protein comprising the sequence as shown in SEQ. ID. NO: 1 and variants thereof that are at least 95% identical to SEQ ID. NO: 2 and specifically binds spectrin.

25. An isolated antibody that specifically binds to a polypeptide having the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 10, or a binding fragment thereof.

26. The antibody of claim 25, that is monoclonal.

27. A hybridoma capable of secreting the antibody of claim 26

28. A method for identifying a compound or agent that binds a CLASP-2 polypeptide comprising:

i) contacting a CLASP-2 polypeptide of claim 19 with the compound or agent under conditions which allow binding of the compound to the CLASP-2 polypeptide to form a complex and

ii) detecting the presence of the complex.

29. A method of detecting a CLASP-2 polypeptide in a sample, comprising:

(a) contacting the sample with an antibody or binding fragment of claim 26 and (b) determining whether a complex has been formed between the antibody and with CLASP-2 polypeptide.

30. A method of detecting a CLASP-2 polypeptide in a sample, comprising:

(a) contacting the sample with a polynucleotide of claim 1 or a polynucleotide that comprises a sequence of at least 12 nucleotides and is complementary to a contiguous sequence of the polynucleotide of section (a) of claim 1, and (b) determining whether a hybridization complex has been formed.

31. A method of detecting a CLASP-2 nucleotide in a sample, comprising:
(a) using a polynucleotide that comprises a sequence of at least 12 nucleotides and is complementary to a contiguous sequence of the polynucleotide of section (a) of claim 1, in an amplification process; and

(b) determining whether a specific amplification product has been formed.

32. A pharmaceutical composition comprising a polynucleotide of claim 1, a polypeptide of claim 18, or an antibody of claim 25 and a pharmaceutically acceptable carrier.

33. A method of inhibiting an immune response in a subject comprising:
(a) interfering with the expression of a CLASP-2 gene;
(b) interfering with the ability of a CLASP-2 protein to bind to another cell;

(c) interfering with the ability of a CLASP-2 protein to bind to another protein.

34. The method of claim 33, wherein the cell is a T cell or a B cell.

35. The method of claim 33 comprising contacting the cell with an effective amount of a polypeptide which comprises the amino acid sequence of SEQ ID NO: 2, 4, 6 or 10 or a fragment thereof.

36. A method of inhibiting an immune response in a subject, comprising administering to the subject a therapeutically effective amount of an antibody which specifically binds a polypeptide having the sequence of SEQ ID NO: 2, 4, 6 or 10.

37. A method of preventing or treating a CLASP-2-mediated disease comprising administering to a subject in need thereof a therapeutically effective amount of a pharmaceutical composition of claim 32.

38. The method claim 37, wherein the CLASP-2-mediated disease is an autoimmune disease.

39. A method of treating an autoimmune disease in a subject caused or exacerbated by increased activity of T_H1 cells consisting of administering a therapeutically effective amount of a pharmaceutical composition of claim 32 to the subject.

1
A

```
2                                                              32
GTT TTA CAC CAT CAC CAA AAC CCA GAA TTT TAT GAT GAG ATT AAA ATA GAG TTG CCC ACT
val leu his his his gln asn pro glu phe tyr asp glu ile lys ile glu leu pro thr

62                                                             92
CAG CTG CAT GAA AAG CAC CAC CTG TTG CTC ACA TTC TTC CAT GTC AGC TGT GAC AAC TCA
gln leu his glu lys his his leu leu leu thr phe phe his val ser cys asp asn ser

122                                                            152
AGT AAA GGA AGC ACG AAG AAG AGG GAT GTC GTT GAA ACC CAA GTT GGC TAC TCC TGG CTT
ser lys gly ser thr lys lys arg asp val val glu thr gln val gly tyr ser trp leu

182                                                            212
CCC CTC CTG AAA GAC GGA AGG GTG GTG ACA AGC GAG CAG CAC ATC CCG GTC TCG GCG AAC
pro leu leu lys asp gly arg val val thr ser glu gln his ile pro val ser ala asn

242                                                            272
CTT CCT TCG GGC TAT CTT GGC TAC CAA GAG CTT GGG ATG GGC AGG CAT TAT GGT CCG GAA
leu pro ser gly tyr leu gly tyr gln glu leu gly met gly arg his tyr gly pro glu

302                                                            332
ATT AAA TGG GTA GAT GGA GGC AAG CCA CTG CTG AAA ATT TCC ACT CAT CTG GTT TCT ACA
ile lys trp val asp gly gly lys pro leu leu lys ile ser thr his leu val ser thr

362                                                            392
GTG TAT ACT CAG GAT CAG CAT TTA CAT AAT TTT TTC CAG TAC TGT CAG AAA ACC GAA TCT
val tyr thr gln asp gln his leu his asn phe phe gln tyr cys gln lys thr glu ser

422                                                            452
GGA GCC CAA GCC TTA GGA AAC GAA CTT GTA AAG TAC CTT AAG AGT CTG CAT GCG ATG GAA
gly ala gln ala leu gly asn glu leu val lys tyr leu lys ser leu his ala met glu

482                                                            512
GGC CAC GTG ATG ATC GCC TTC TTG CCC ACT ATC CTA AAC CAG CTG TTC CGA GTC CTC ACC
gly his val met ile ala phe leu pro thr ile leu asn gln leu phe arg val leu thr

542                                                            572
AGA GCC ACA CAG GAA GAA GTC GCG GTT AAC GTG ACT CGG GTC ATT ATT CAT GTG GTT GCC
arg ala thr gln glu glu val ala val asn val thr arg val ile ile his val val ala

602                                                            632
CAG TGC CAT GAG GAA GGA TTG GAG AGC CAC TTG AGG TCA TAT GTT AAG TAC GCG TAT AAG
gln cys his glu glu gly leu glu ser his leu arg ser tyr val lys tyr ala tyr lys

662                                                            692
GCT GAG CCA TAT GTT GCC TCT GAA TAC AAG ACA GTG CAT GAA GAA CTG ACC AAA TCC ATG
ala glu pro tyr val ala ser glu tyr lys thr val his glu glu leu thr lys ser met
```

FIG. 1

```
722                                                 752
ACC ACG ATT CTC AAG CCT TCT GCC GAT TTC CTC ACC AGC AAC AAA CTA CTG AGG TAC TCA
thr thr ile leu lys pro ser ala asp phe leu thr ser asn lys leu leu arg tyr ser

782                                                 812
TGG TTT TTC TTT GAT GTA CTG ATC AAA TCT ATG GCT CAG CAT TTG ATA GAG AAC TCC AAA
trp phe phe phe asp val leu ile lys ser met ala gln his leu ile glu asn ser lys

842              |Cadherin Cleavage|                872
GTT AAG TTG CTG CGA AAC CAG AGA TTT CCT GCA TCC TAT CAT CAT GCA GCG GAA ACC GTT
val lys leu leu arg asn gln arg phe pro ala ser tyr his his ala ala glu thr val

902                                                 932
GTA AAT ATG CTG ATG CCA CAC ATC ACT CAG AAG TTT GGA GAT AAT CCA GAG GCA TCT AAG
val asn met leu met pro his ile thr gln lys phe gly asp asn pro glu ala ser lys

962                                                 992
AAC GCG AAT CAT AGC CTT GCT GTC TTC ATC AAG AGA TGT TTC ACC TTC ATG GAC AGG GGC
asn ala asn his ser leu ala val phe ile lys arg cys phe thr phe met asp arg gly

1022                                                1052
TTT GTC TTC AAG CAG ATC AAC AAC TAC ATT AGC TGT TTT GCT CCT GGA GAC CCA AAG ACC
phe val phe lys gln ile asn asn tyr ile ser cys phe ala pro gly asp pro lys thr

1082                                                1112
CTC TTT GAA TAC AAG TTT GAA TTT CTC CGT GTA GTG TGC AAC CAT GAA CAT TAT ATT CCG
leu phe glu tyr lys phe glu phe leu arg val val cys asn his glu his tyr ile pro

1142                                                1172
TTG AAC TTA CCA ATG CCA TTT GGA AAA GGC AGG ATT CAA AGA TAC CAA GAC CTC CAG CTT
leu asn leu pro met pro phe gly lys gly arg ile gln arg tyr gln asp leu gln leu

1202                                                1232                    |Cadherin EC
GAC TAC TCA TTA ACA GAT GAG TTC TGC AGA AAC CAC TTC TTG GTG GGA CTG TTA CTG AGG
asp tyr ser leu thr asp glu phe cys arg asn his phe leu val gly leu leu leu arg

xxxx|                                               1292
GAG GTG GGG ACA GCC CTC CAG GAG TTC CGG GAG GTC CGT CTG ATC GCC ATC AGT GTG CTC
glu val gly thr ala leu gln glu phe arg glu val arg leu ile ala ile ser val leu

1322                                                1352
AAG AAC CTG CTG ATA AAG CAT TCT TTT GAT GAC AGA TAT GCT TCA AGG AGC CAT CAG GCA
lys asn leu leu ile lys his ser phe asp asp arg tyr ala ser arg ser his gln ala

1382                                                1412/471
AGG ATA GCC ACC CTC TAC CTG CCT CTG TTT GGT CTG CTG ATT GAA AAC GTC CAG CGG ATC
arg ile ala thr leu tyr leu pro leu phe gly leu leu ile glu asn val gln arg ile

1442                                                1472
AAT GTG AGG GAT GTG TCA CCC TTC CCT GTG AAC GCG GGC ATG ACC GTG AAG GAT GAA TCC
asn val arg asp val ser pro phe pro val asn ala gly met thr val lys asp glu ser
```

FIG. 1 (cont.)

```
1502                                                1532
CTG GCT CTA CCA GCT GTG AAT CCG CTG GTG ACG CCG CAG AAG GGA AGC ACC CTG GAC AAC
leu ala leu pro ala val asn pro leu val thr pro gln lys gly ser thr leu asp asn

1562                                                1592
AGC CTG CAC AAC GAC CTG CTG GGC GCC ATC TCC GGC ATT GCT TCT CCA TAT ACA ACC TCA
ser leu his lys asp leu leu gly ala ile ser gly ile ala ser pro tyr thr thr ser

1622                                                1652
ACT CCA AAC ATC AAC AGT GTG AGA AAT GCT GAT TCG AGA GGA TCT CTC ATA AGC ACA GAT
thr pro asn ile asn ser val arg asn ala asp ser arg gly ser leu ile ser thr asp

1682                                                1712
TCG GGT AAC AGC CTT CCA GAA AGG AAT AGT GAG AAG AGC AAT TCC CTG GAT AAG CAC CAA
ser gly asn ser leu pro glu arg asn ser glu lys ser asn ser leu asp lys his gln

1742                                                1772
CAA AGT AGC ACA TTG GGA AAT TCC GTG GTT CGC TGT GAT AAA CTT GAC CAG TCT GAG ATT
gln ser ser thr leu gly asn ser val val arg cys asp lys leu asp gln ser glu ile

1802                                                1832
AAG AGC CTA CTG ATG TGT TTC CTC TAC ATC TTA AAG AGC ATG TCT GAT GAT GCT TTG TTT
lys ser leu leu met cys phe leu tyr ile leu lys ser met ser asp asp ala leu phe

1862                                                1892
ACA TAT TGG AAC AAG GCT TCA ACA TCT GAA CTT ATG GAT TTT TTT ACA ATA TCT GAA GTC
thr tyr trp asn lys ala ser thr ser glu leu met asp phe phe thr ile ser glu val

1922                                                1952
TGC CTG CAC CAG TTC CAG TAC ATG GGG AAG CGA TAC ATA GCC AGG AAC CAG GAG GGG TTG
cys leu his gln phe gln tyr met gly lys arg tyr ile ala arg asn gln glu gly leu

1982                                                2012
GGA CCC ATA GTT CAT GAT CGA AAG TCT CAG ACA TTG CCT GTT TCC CGT AAC AGA ACA GGA
gly pro ile val his asp arg lys ser gln thr leu pro val ser arg asn arg thr gly

2042                                                2072
ATG ATG CAT GCC AGA TTG CAG CAG CTG GGC AGC CTG GAT AAC TCT CTC ACT TTT AAC CAC
met met his ala arg leu gln gln leu gly ser leu asp asn ser leu thr phe asn his

2102                                                2132
AGC TAT GGC CAC TCG GAC GCA GAT GTT CTG CAC CAG TCA TTA CTT GAA GCC AAC ATT GCT
ser tyr gly his ser asp ala asp val leu his gln ser leu leu glu ala asn ile ala

2162                                                2192
ACT GAG GTT TGC CTG ACA GCT CTG GAC ACG CTT TCT CTA TTT ACA TTG GCG TTT AAG AAC
thr glu val cys leu thr ala leu asp thr leu ser leu phe thr leu ala phe lys asn

2222                                                2252
CAG CTC CTG GCC GAC CAT GGA CAT AAT CCT CTC ATG AAA AAA GTT TTT GAT GTC TAC CTG
gln leu leu ala asp his gly his asn pro leu met lys lys val phe asp val tyr leu
```

FIG. 1 (cont.)

```
2282                                                2312
TGT TTT CTT CAA AAA CAT CAG TCT GAA ACG GCT TTA AAA AAT GTC TTC ACT GCC TTA AGG
cys phe leu gln lys his gln ser glu thr ala leu lys asn val phe thr ala leu arg

2342                                                2372
TCC TTA ATT TAT AAG TTT CCC TCA ACA TTC TAT GAA GGG AGA GCG GAC ATG TGT GCG GCT
ser leu ile tyr lys phe pro ser thr phe tyr glu gly arg ala asp met cys ala ala

2402                                                2432
CTG TGT TAC GAG ATT CTC AAG TGC TGT AAC TCC AAG CTG AGC TCC ATC AGG ACG GAG GCC
leu cys tyr glu ile leu lys cys cys asn ser lys leu ser ser ile arg thr glu ala

2462                                                2492
TCC CAG CTG CTC TAC TTC CTG ATG AGG AAC AAC TTT GAT TAC ACT GGA AAG AAG TCC TTT
ser gln leu leu tyr phe leu met arg asn asn phe asp tyr thr gly lys lys ser phe

2522                                                2552
GTC CGG ACA CAT TTG CAA GTC ATC ATA TCT GTC AGC CAG CTG ATA GCA GAC GTT GTT GGC
val arg thr his leu gln val ile ile ser val ser gln leu ile ala asp val val gly

2582                                                2612
ATT GGG GAA ACC AGA TTC CAG CAG TCC CTG TCC ATC ATC AAC AAC TGT GCC AAC AGT GAC
ile gly glu thr arg phe gln gln ser leu ser ile ile asn asn cys ala asn ser asp

2642                                                2672
CGG CTT ATT AAG CAC ACC AGC TTC TCC TCT GAT GTG AAG GAC TTA ACC AAA AGG ATA CGC
arg leu ile lys his thr ser phe ser ser asp val lys asp leu thr lys arg ile arg

2702                                                2732
ACG GTG CTA ATG GCC ACC GCC CAG ATG AAG GAG CAT GAG AAC GAC CCA GAG ATG CTG GTG
thr val leu met ala thr ala gln met lys glu his glu asn asp pro glu met leu val

2762                                                2792
GAC CTC CAG TAC AGC CTG GCC AAA TCC TAT GCC AGC ACG CCC GAG CTC AGG AAG ACG TGG
asp leu gln tyr ser leu ala lys ser tyr ala ser thr pro glu leu arg lys thr trp

2822                                                2852          |xxxxxxxxxxxxxxx Predicted
CTC GAC AGC ATG GCC AGG ATC CAT GTC AAA AAT GGC GAT CTC TCA GAG GCA GCA ATG TGC
leu asp ser met ala arg ile his val lys asn gly asp leu ser glu ala ala met cys

Transmembrane Domain xxxxxxxxxxxxxxxxxxxxxxxxxxxx|
TAT GTC CAC GTA ACA GCC CTA GTG GCA GAA TAT CTC ACA CGG AAA GGC GTG TTT AGA CAA
tyr val his val thr ala leu val ala glu tyr leu thr arg lys gly val phe arg gln

2942                                                2972
GGA TGC ACC GCC TTC AGG GTC ATT ACC CCA AAC ATC GAC GAG GAG GCC TCC ATG ATG GAA
gly cys thr ala phe arg val ile thr pro asn ile asp glu glu ala ser met met glu

3002                                                3032
GAC GTG GGG ATG CAG GAT GTC CAT TTC AAC GAG GAT GTG CTG ATG GAG CTC CTT GAG CAG
asp val gly met gln asp val his phe asn glu asp val leu met glu leu leu glu gln

3062                                                3092
TGC GCA GAT GGA CTC TGG AAA GCC GAG CGC TAC GAG CTC ATC GCC GAC ATC TAC AAA CTT
cys ala asp gly leu trp lys ala glu arg tyr glu leu ile ala asp ile tyr lys leu
```

FIG. 1 (cont.)

```
3122                                                3155
ATC ATC CCC ATT TAT GAG AAG CGG AGG GAT TTC TTT GAA GAT GAA GAT GGA AAG GAG TAT
ile ile pro ile tyr glu lys arg arg asp phe phe glu asp glu asp gly lys glu tyr

3182                                                3212
ATT TAC AAG GAA CCC AAA CTC ACA CCG CTG TCG GAA ATT TCT CAG AGA CTC CTT AAA CTG
ile tyr lys glu pro lys leu thr pro leu ser glu ile ser gln arg leu leu lys leu

3242                                                3272
TAC TCG GAT AAA TTT GGT TCT GAA AAT GTC AAA ATG ATA CAG GAT TCT GGC AAG GTC AAC
tyr ser asp lys phe gly ser glu asn val lys met ile gln asp ser gly lys val asn

3302                                                3332
CCT AAG GAT CTG GAT TCT AAG TAT GCA TAC ATC CAG GTG ACT CAC GTC ATC CCC TTC TTT
pro lys asp leu asp ser lys tyr ala tyr ile gln val thr his val ile pro phe phe

3362                                                3392
GAC GAA AAA GAG TTG CAA GAA AGG AAA ACA GAG TTT GAG AGA TCC CAC AAC ATC CGC CGC
asp glu lys glu leu gln glu arg lys thr glu phe glu arg ser his asn ile arg arg

3422                                                3452
TTC ATG TTT GAG ATG CCA TTT ACG CAG ACC GGG AAG AGG CAG GGC GGG GTG GAA GAG CAG
phe met phe glu met pro phe thr gln thr gly lys arg gln gly gly val glu glu gln

3482                                                3512
TGC AAA CGG CGC ACC ATC CTG ACA GCC ATA CAC TGC TTC CCT TAT GTG AAG AAG CGC ATC
cys lys arg arg thr ile leu thr ala ile his cys phe pro tyr val lys lys arg ile

3542                                                3572 |xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
CCT GTC ATG TAC CAG CAC CAC ACT GAC CTG AAC CCC ATC GAG GTG GCC ATT GAC GAG ATG
pro val met tyr gln his his thr asp leu asn pro ile glu val ala ile asp glu met

3602 xxx Coiled-Coil 1 xxxxxxxxxxxxx 3632 xxxx Coiled-Coil 1 xxxxxxxxxxxxx
AGT AAG AAG GTG GCG GAG CTC CGG CAG CTG TGC TCC TCG GCC GAG GTG GAC ATG ATC AAA
ser lys lys val ala glu leu arg gln leu cys ser ser ala glu val asp met ile lys

3662 xxxxxxxxxxxxxxxxxxxxxxxxx|                  3692
CTG CAG CTC AAA CTC CAG GGC AGC GTG AGT GTT CAG GTC AAT GCT GGC CCA CTA GCA TAT
leu gln leu lys leu gln gly ser val ser val gln val asn ala gly pro leu ala tyr

3722                                                3752
GCG CGA GCT TTC TTA GAT GAT ACA AAC ACA AAG CGA TAT CCT GAC AAT AAA GTG AAG CTG
ala arg ala phe leu asp asp thr asn thr lys arg tyr pro asp asn lys val lys leu

3782                                                3812                    |xxxxxxxxxxxxxxxxxx
CTT AAG GAA GTT TTC AGG CAA TTT GTG GAA GCT TGC GGT CAA GCC TTA GCG GTA AAC GAA
leu lys glu val phe arg gln phe val glu ala cys gly gln ala leu ala val asn glu
```

FIG. 1 (cont.)

```
3842 xxx  Coiled-Coil 2  xxxxxxxxxxxxx  3872 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
CGT CTG ATT AAA GAA GAC CAG CTC GAG TAT CAG GAA GAA ATG AAA GCC AAC TAC AGG GAA
arg leu ile lys glu asp gln leu glu tyr gln glu glu met lys ala asn tyr arg glu

3902 xxx  Coiled-Coil 2  xxxxxxxxxxxxx  3932 xxxxx|
ATG GCG AAG GAG CTT TCT GAA ATC ATG CAT GAG CAG ATC TGC CCC CTG GAG GAG AAG ACG
met ala lys glu leu ser glu ile met his glu gln ile cys pro leu glu glu lys thr

3962                                    3992
AGC GTC TTA CCG AAT TCC CTT CAC ATC TTC AAC GCC ATC AGT GGG ACT CCA ACA AGC ACA
ser val leu pro asn ser leu his ile phe asn ala ile ser gly thr pro thr ser thr

4022                                |xxxx PBM xxxx|
ATG GTT CAC GGG ATG ACC AGC TCG TCT TCG GTC GTG TGA TTA CAT CTC ATG GCC CGT GTG
met val his gly met thr ser ser ser ser val val STP

4082                                        4112
TGG GGA CTT GCT TTG TCA TTT GCA AAC TCA GGA TGC TTT CCA AAG CCA ATC ACT GGG GAG

4142                                        4172
ACC GAG CAC AGG GAG GAC CAA GGG GAA GGG GAG AGA AAG GAA ATA AAG AAC AAC GTT ATT

4202                                        4232
TCT TAA CAG ACT TTC TAT AGG AGT TGT AAG AAG GTG CAC ATA TTT TTT TAA ATC TCA CTG

4262                                        4292
GCA ATA TTC AAA GTT TTC ATT GTG TCT TAA CAA AGG TGT GGT AGA CAC TCT TGA GCT GGA

4322                                        4352
CTT AGA TTT TAT TCT TCC TTG CAG AGT AGT GTT AGA ATA GAT GGC CTA CAG AAA AAA AAG

4382                                        4412
GTT CTG GGA TCT ACA TGG CAG GGA GGG CTG CAC TGA CAT TGA TGC CTG GGG GAC CTT TTG

4442                                        4472
CCT CGA CTC GTG CCG GAA ATC TGA TCG TAA TCA GGG TAC AGA ACT TAC TAG TTT TGT CTA

4502                                        4532
GGA GTA TGT TGT ATG ACT AGG ATT TGT GCT ATT ATC TCA TTC AAC AAC ATA GAG CAA GAA

4562                                        4592
TAG TGA GCT AAC TGA GCT AGA CAC TCA ATT AAT CCG CTA CTG GCT TCA AGT CAG AAC TTT

4622                                        4652
GTC ATT AAT CAT CGA CTC CGG GAC GGT CAT ATA TGT ATT ACA TTT CTA CAT TTT TAA TAC

4682                                        4712
TCA CAT GGG CTT ATG CAT TAA GTT TAA TTG TGA TAA ATT TGT GCT GGT CCA GTA TAT GCA

4742                                        4772
ATA CAC TTT AAT GGT TTA TTC TTG TCA TAA AAA TGT GCA ATA TGG AGA TGT ATA CAA GTC

4802
TTT ACT
```

FIG. 1 (cont.)

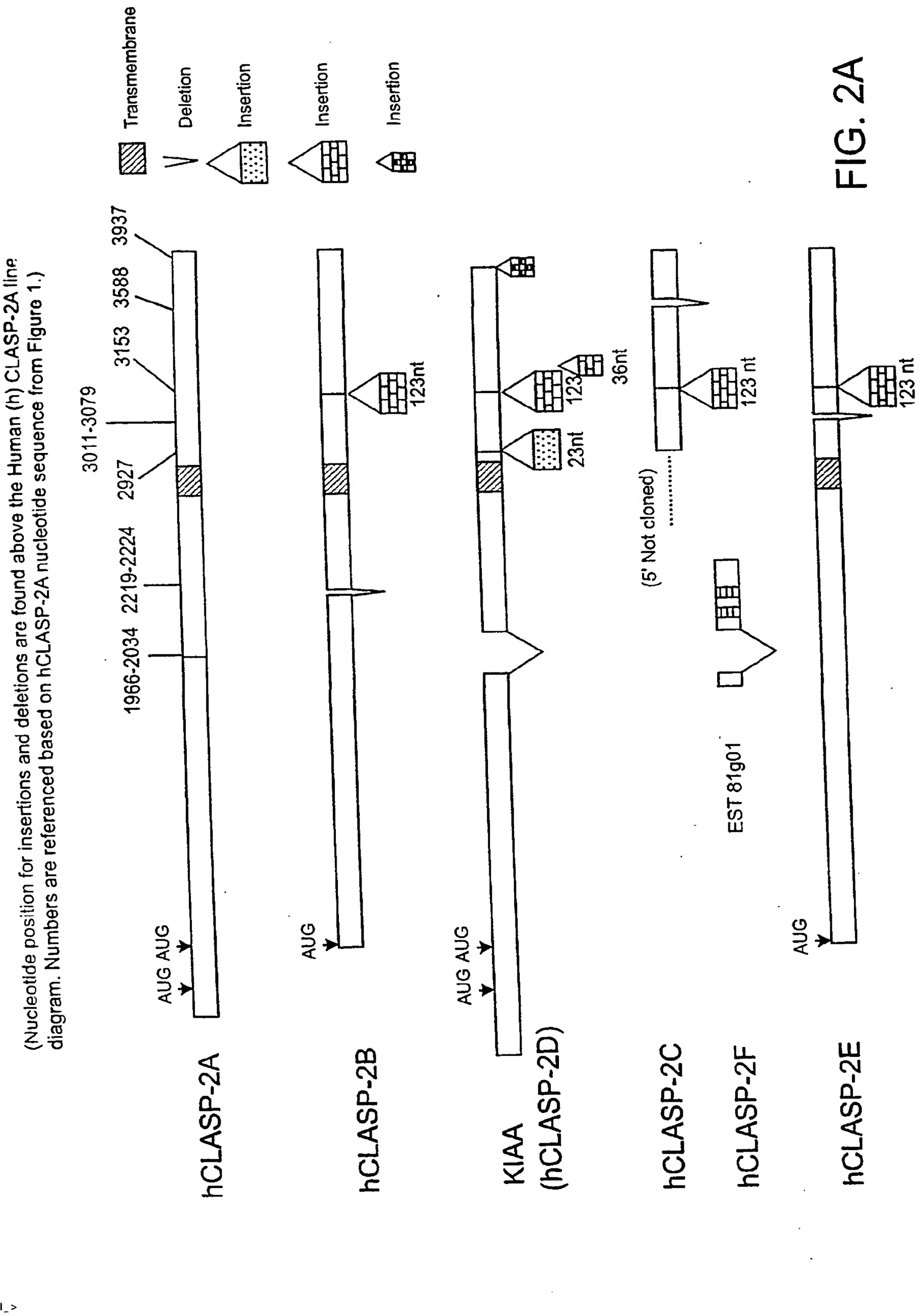

Transmembrane
Deletion
Insertion
Insertion
Insertion
(Nucleotide position for insertions and deletions are found above the Human (h) CLASP-2A line diagram. Numbers are referenced based on hCLASP-2A nucleotide sequence from Figure 1.)
3011-3079
3153
3588
3937
1966-2034
2219-2224
2927
123nt
23nt
36nt
(5' Not cloned)
123 nt
AUG AUG
hCLASP-2A
AUG
hCLASP-2B
AUG AUG
KIAA
(hCLASP-2D)
hCLASP-2C
EST 81g01
hCLASP-2F
AUG
hCLASP-2E

FIG. 2A

1
A

```
2                                                   32
GTT TTA CAC CAT CAC CAA AAC CCA GAA TTT TAT GAT GAG ATT AAA ATA GAG TTG CCC ACT
val leu his his his gln asn pro glu phe tyr asp glu ile lys ile glu leu pro thr

62                                                  92
CAG CTG CAT GAA AAG CAC CAC CTG TTG CTC ACA TTC TTC CAT GTC AGC TGT GAC AAC TCA
gln leu his glu lys his his leu leu leu thr phe phe his val ser cys asp asn ser

122                                                 152
AGT AAA GGA AGC ACG AAG AAG AGG GAT GTC GTT GAA ACC CAA GTT GGC TAC TCC TGG CTT
ser lys gly ser thr lys lys arg asp val val glu thr gln val gly tyr ser trp leu

182                                                 212
CCC CTC CTG AAA GAC GGA AGG GTG GTG ACA AGC GAG CAG CAC ATC CCG GTC TCG GCG AAC
pro leu leu lys asp gly arg val val thr ser glu gln his ile pro val ser ala asn

242                                                 272
CTT CCT TCG GGC TAT CTT GGC TAC CAA GAG CTT GGG ATG GGC AGG CAT TAT GGT CCG GAA
leu pro ser gly tyr leu gly tyr gln glu leu gly met gly arg his tyr gly pro glu

302                                                 332
ATT AAA TGG GTA GAT GGA GGC AAG CCA CTG CTG AAA ATT TCC ACT CAT CTG GTT TCT ACA
ile lys trp val asp gly gly lys pro leu leu lys ile ser thr his leu val ser thr

362                                                 392
GTG TAT ACT CAG GAT CAG CAT TTA CAT AAT TTT TTC CAG TAC TGT CAG AAA ACC GAA TCT
val tyr thr gln asp gln his leu his asn phe phe gln tyr cys gln lys thr glu ser

422                                                 452
GGA GCC CAA GCC TTA GGA AAC GAA CTT GTA AAG TAC CTT AAG AGT CTG CAT GCG ATG GAA
gly ala gln ala leu gly asn glu leu val lys tyr leu lys ser leu his ala met glu

482                                                 512
GGC CAC GTG ATG ATC GCC TTC TTG CCC ACT ATC CTA AAC CAG CTG TTC CGA GTC CTC ACC
gly his val met ile ala phe leu pro thr ile leu asn gln leu phe arg val leu thr

542                                                 572
AGA GCC ACA CAG GAA GAA GTC GCG GTT AAC GTG ACT CGG GTC ATT ATT CAT GTG GTT GCC
arg ala thr gln glu glu val ala val asn val thr arg val ile ile his val val ala

602                                                 632
CAG TGC CAT GAG GAA GGA TTG GAG AGC CAC TTG AGG TCA TAT GTT AAG TAC GCG TAT AAG
gln cys his glu glu gly leu glu ser his leu arg ser tyr val lys tyr ala tyr lys

662                                                 692
GCT GAG CCA TAT GTT GCC TCT GAA TAC AAG ACA GTG CAT GAA GAA CTG ACC AAA TCC ATG
ala glu pro tyr val ala ser glu tyr lys thr val his glu glu leu thr lys ser met
```

FIG. 2B

```
722                                                 752
ACC ACG ATT CTC AAG CCT TCT GCC GAT TTC CTC ACC AGC AAC AAA CTA CTG AGG TAC TCA
thr thr ile leu lys pro ser ala asp phe leu thr ser asn lys leu leu arg tyr ser

782                                                 812
TGG TTT TTC TTT GAT GTA CTG ATC AAA TCT ATG GCT CAG CAT TTG ATA GAG AAC TCC AAA
trp phe phe phe asp val leu ile lys ser met ala gln his leu ile glu asn ser lys

842              |Cadherin Cleavage|                872
GTT AAG TTG CTG CGA AAC CAG AGA TTT CCT GCA TCC TAT CAT CAT GCA GCG GAA ACC GTT
val lys leu leu arg asn gln arg phe pro ala ser tyr his his ala ala glu thr val

902                                                 932
GTA AAT ATG CTG ATG CCA CAC ATC ACT CAG AAG TTT GGA GAT AAT CCA GAG GCA TCT AAG
val asn met leu met pro his ile thr gln lys phe gly asp asn pro glu ala ser lys

962                                                 992
AAC GCG AAT CAT AGC CTT GCT GTC TTC ATC AAG AGA TGT TTC ACC TTC ATG GAC AGG GGC
asn ala asn his ser leu ala val phe ile lys arg cys phe thr phe met asp arg gly

1022                                               1052
TTT GTC TTC AAG CAG ATC AAC AAC TAC ATT AGC TGT TTT GCT CCT GGA GAC CCA AAG ACC
phe val phe lys gln ile asn asn tyr ile ser cys phe ala pro gly asp pro lys thr

1082                                               1112
CTC TTT GAA TAC AAG TTT GAA TTT CTC CGT GTA GTG TGC AAC CAT GAA CAT TAT ATT CCG
leu phe glu tyr lys phe glu phe leu arg val val cys asn his glu his tyr ile pro

1142                                               1172
TTG AAC TTA CCA ATG CCA TTT GGA AAA GGC AGG ATT CAA AGA TAC CAA GAC CTC CAG CTT
leu asn leu pro met pro phe gly lys gly arg ile gln arg tyr gln asp leu gln leu

1202                                               1232                    |Cadherin EC
GAC TAC TCA TTA ACA GAT GAG TTC TGC AGA AAC CAC TTC TTG GTG GGA CTG TTA CTG AGG
asp tyr ser leu thr asp glu phe cys arg asn his phe leu val gly leu leu leu arg

xxxx|                                              1292
GAG GTG GGG ACA GCC CTC CAG GAG TTC CGG GAG GTC CGT CTG ATC GCC ATC AGT GTG CTC
glu val gly thr ala leu gln glu phe arg glu val arg leu ile ala ile ser val leu

1322                                               1352
AAG AAC CTG CTG ATA AAG CAT TCT TTT GAT GAC AGA TAT GCT TCA AGG AGC CAT CAG GCA
lys asn leu leu ile lys his ser phe asp asp arg tyr ala ser arg ser his gln ala

1382                                               1412
AGG ATA GCC ACC CTC TAC CTG CCT CTG TTT GGT CTG CTG ATT GAA AAC GTC CAG CGG ATC
arg ile ala thr leu tyr leu pro leu phe gly leu leu ile glu asn val gln arg ile

1442                                               1472
AAT GTG AGG GAT GTG TCA CCC TTC CCT GTG AAC GCG GGC ATG ACC GTG AAG GAT GAA TCC
asn val arg asp val ser pro phe pro val asn ala gly met thr val lys asp glu ser
```

FIG. 2B (cont.)

```
1502                                                1532
CTG GCT CTA CCA GCT GTG AAT CCG CTG GTG ACG CCG CAG AAG GGA AGC ACC CTG GAC AAC
leu ala leu pro ala val asn pro leu val thr pro gln lys gly ser thr leu asp asn

1562                                                1592
AGC CTG CAC AAG GAC CTG CTG GGC GCC ATC TCC GGC ATT GCT TCT CCA TAT ACA ACC TCA
ser leu his lys asp leu leu gly ala ile ser gly ile ala ser pro tyr thr thr ser

1622                                                1652
ACT CCA AAC ATC AAC AGT GTG AGA AAT GCT GAT TCG AGA GGA TCT CTC ATA AGC ACA GAT
thr pro asn ile asn ser val arg asn ala asp ser arg gly ser leu ile ser thr asp

1682                                                1712
TCG GGT AAC AGC CTT CCA GAA AGG AAT AGT GAG AAG AGC AAT TCC CTG GAT AAG CAC CAA
ser gly asn ser leu pro glu arg asn ser glu lys ser asn ser leu asp lys his gln

1742                                                1772
CAA AGT AGC ACA TTG GGA AAT TCC GTG GTT CGC TGT GAT AAA CTT GAC CAG TCT GAG ATT
gln ser ser thr leu gly asn ser val val arg cys asp lys leu asp gln ser glu ile

1802                                                1832
AAG AGC CTA CTG ATG TGT TTC CTC TAC ATC TTA AAG AGC ATG TCT GAT GAT GCT TTG TTT
lys ser leu leu met cys phe leu tyr ile leu lys ser met ser asp asp ala leu phe

1862                                                1892
ACA TAT TGG AAC AAG GCT TCA ACA TCT GAA CTT ATG GAT TTT TTT ACA ATA TCT GAA GTC
thr tyr trp asn lys ala ser thr ser glu leu met asp phe phe thr ile ser glu val

1922                                                1952                    |xxxxxxxxxxxxxxxxxxxxx
TGC CTG CAC CAG TTC CAG TAC ATG GGG AAG CGA TAC ATA GCC AGG AAC CAG GAG GGG TTG
cys leu his gln phe gln tyr met gly lys arg tyr ile ala arg asn gln glu gly leu

1982 xxxxxxxxxx  deleted in CLASP-2D(KIAA1058) xxxxxxxxxxxxxxxxxxxxxx|
GGA CCC ATA GTT CAT GAT CGA AAG TCT CAG ACA TTG CCT GTT TCC CGT AAC AGA ACA GGA
gly pro ile val his asp arg lys ser gln thr leu pro val ser arg asn arg thr gly

2042                                                2072
ATG ATG CAT GCC AGA TTG CAG CAG CTG GGC AGC CTG GAT AAC TCT CTC ACT TTT AAC CAC
met met his ala arg leu gln gln leu gly ser leu asp asn ser leu thr phe asn his

2102                                                2132
AGC TAT GGC CAC TCG GAC GCA GAT GTT CTG CAC CAG TCA TTA CTT GAA GCC AAC ATT GCT
ser tyr gly his ser asp ala asp val leu his gln ser leu leu glu ala asn ile ala

                                                                                    Deleted
2162                                                2192                             |xxx
ACT GAG GTT TGC CTG ACA GCT CTG GAC ACG CTT TCT CTA TTT ACA TTG GCG TTT AAG AAC
thr glu val cys leu thr ala leu asp thr leu ser leu phe thr leu ala phe lys asn

in HC2B
xxx|                                                2252
CAG CTC CTG GCC GAC CAT GGA CAT AAT CCT CTC ATG AAA AAA GTT TTT GAT GTC TAC CTG
gln leu leu ala asp his gly his asn pro leu met lys lys val phe asp val tyr leu
```

FIG. 2B (cont.)

```
2282                                                2312
TGT TTT CTT CAA AAA CAT CAG TCT GAA ACG GCT TTA AAA AAT GTC TTC ACT GCC TTA AGG
cys phe leu gln lys his gln ser glu thr ala leu lys asn val phe thr ala leu arg

2342                                                2372
TCC TTA ATT TAT AAG TTT CCC TCA ACA TTC TAT GAA GGG AGA GCG GAC ATG TGT GCG GCT
ser leu ile tyr lys phe pro ser thr phe tyr glu gly arg ala asp met cys ala ala

2402                                                2432
CTG TGT TAC GAG ATT CTC AAG TGC TGT AAC TCC AAG CTG AGC TCC ATC AGG ACG GAG GCC
leu cys tyr glu ile leu lys cys cys asn ser lys leu ser ser ile arg thr glu ala

2462                                                2492
TCC CAG CTG CTC TAC TTC CTG ATG AGG AAC AAC TTT GAT TAC ACT GGA AAG AAG TCC TTT
ser gln leu leu tyr phe leu met arg asn asn phe asp tyr thr gly lys lys ser phe

2522                                                2552
GTC CGG ACA CAT TTG CAA GTC ATC ATA TCT GTC AGC CAG CTG ATA GCA GAC GTT GTT GGC
val arg thr his leu gln val ile ile ser val ser gln leu ile ala asp val val gly

2582                                                2612
ATT GGG GAA ACC AGA TTC CAG CAG TCC CTG TCC ATC ATC AAC AAC TGT GCC AAC AGT GAC
ile gly glu thr arg phe gln gln ser leu ser ile ile asn asn cys ala asn ser asp

2642                                                2672
CGG CTT ATT AAG CAC ACC AGC TTC TCC TCT GAT GTG AAG GAC TTA ACC AAA AGG ATA CGC
arg leu ile lys his thr ser phe ser ser asp val lys asp leu thr lys arg ile arg

2702                                                2732
ACG GTG CTA ATG GCC ACC GCC CAG ATG AAG GAG CAT GAG AAC GAC CCA GAG ATG CTG GTG
thr val leu met ala thr ala gln met lys glu his glu asn asp pro glu met leu val

2762                                                2792
GAC CTC CAG TAC AGC CTG GCC AAA TCC TAT GCC AGC ACG CCC GAG CTC AGG AAG ACG TGG
asp leu gln tyr ser leu ala lys ser tyr ala ser thr pro glu leu arg lys thr trp

2822                                    2852        |xxxxxxxxxxxxxxxx Predicted
CTC GAC AGC ATG GCC AGG ATC CAT GTC AAA AAT GGC GAT CTC TCA GAG GCA GCA ATG TGC
leu asp ser met ala arg ile his val lys asn gly asp leu ser glu ala ala met cys

    [Additional and differential exon usage found at position 2927 consisting
    of 69 nucleotides.  This entire sequence is found in Human CLASP-2D
    (KIAA1058) and not other isoforms of CLASP-2. It has a sequence of:
    AAGCAGTCCAGTGGGAGCCGCCCCTTCTCCCCCACAGCCATAGCGCCTGCCTGAGGAGGAGCCGGGGAG]

Transmembrane Domain xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx|
TAT GTC CAC GTA ACA GCC CTA GTG GCA GAA TAT CTC ACA CGG AAA GGC GTG TTT AGA CAA
tyr val his val thr ala leu val ala glu tyr leu thr arg lys gly val phe arg gln

2942                                                2972
GGA TGC ACC GCC TTC AGG GTC ATT ACC CCA AAC ATC GAC GAG GAG GCC TCC ATG ATG GAA
gly cys thr ala phe arg val ile thr pro asn ile asp glu glu ala ser met met glu
```

FIG. 2B (cont.)

3002 |xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx Sequence deleted in CLASP-2E xxxxx
GAC GTG GGG ATG CAG GAT GTC CAT TTC AAC GAG GAT GTG CTG ATG GAG CTC CTT GAG CAG
asp val gly met gln asp val his phe asn glu asp val leu met glu leu leu glu gln

3062 xxxxxxxxxxxxxxxxxx| 3092
TGC GCA GAT GGA CTC TGG AAA GCC GAG CGC TAC GAG CTC ATC GCC GAC ATC TAC AAA CTT
cys ala asp gly leu trp lys ala glu arg tyr glu leu ile ala asp ile tyr lys leu

[Additional and differential exon usage found at position 3153. The entire sequence below is found in Human CLASP-2D. Underlined sequence is found in Human CLASP-2B, 2C and 2E.
TGAGAGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGACCGAGGTCAT
GCACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTCTTCGGGCAGGCAGCGCAATACCAGTTT
ACAGACAGTGAAACAGATGTGGAGGGATT]

3122 3155
ATC ATC CCC ATT TAT GAG AAG CGG AGG GAT TTC TTT GAA GAT GAA GAT GGA AAG GAG TAT
ile ile pro ile tyr glu lys arg arg asp phe phe glu asp glu asp gly lys glu tyr

3182 3212
ATT TAC AAG GAA CCC AAA CTC ACA CCG CTG TCG GAA ATT TCT CAG AGA CTC CTT AAA CTG
ile tyr lys glu pro lys leu thr pro leu ser glu ile ser gln arg leu leu lys leu

3242 3272
TAC TCG GAT AAA TTT GGT TCT GAA AAT GTC AAA ATG ATA CAG GAT TCT GGC AAG GTC AAC
tyr ser asp lys phe gly ser glu asn val lys met ile gln asp ser gly lys val asn

3302 3332
CCT AAG GAT CTG GAT TCT AAG TAT GCA TAC ATC CAG GTG ACT CAC GTC ATC CCC TTC TTT
pro lys asp leu asp ser lys tyr ala tyr ile gln val thr his val ile pro phe phe

3362 3392
GAC GAA AAA GAG TTG CAA GAA AGG AAA ACA GAG TTT GAG AGA TCC CAC AAC ATC CGC CGC
asp glu lys glu leu gln glu arg lys thr glu phe glu arg ser his asn ile arg arg

3422 3452
TTC ATG TTT GAG ATG CCA TTT ACG CAG ACC GGG AAG AGG CAG GGC GGG GTG GAA GAG CAG
phe met phe glu met pro phe thr gln thr gly lys arg gln gly gly val glu glu gln

3482 3512
TGC AAA CGG CGC ACC ATC CTG ACA GCC ATA CAC TGC TTC CCT TAT GTG AAG AAG CGC ATC
cys lys arg arg thr ile leu thr ala ile his cys phe pro tyr val lys lys arg ile

Two nucleotide deletion (nts 3586 and 3587) found in Human CLASP-2C

3542 3572 |xxx|
CCT GTC ATG TAC CAG CAC CAC ACT GAC CTG AAC CCC ATC GAG GTG GCC ATT GAC GAG ATG
pro val met tyr gln his his thr asp leu asn pro ile glu val ala ile asp glu met

FIG. 2B (cont.)

```
3602                                                    3632
AGT AAG AAG GTG GCG GAG CTC CGG CAG CTG TGC TCC TCG GCC GAG GTG GAC ATG ATC AAA
ser lys lys val ala glu leu arg gln leu cys ser ser ala glu val asp met ile lys

3662                                                    3692
CTG CAG CTC AAA CTC CAG GGC AGC GTG AGT GTT CAG GTC AAT GCT GGC CCA CTA GCA TAT
leu gln leu lys leu gln gly ser val ser val gln val asn ala gly pro leu ala tyr

3722                                                    3752
GCG CGA GCT TTC TTA GAT GAT ACA AAC ACA AAG CGA TAT CCT GAC AAT AAA GTG AAG CTG
ala arg ala phe leu asp asp thr asn thr lys arg tyr pro asp asn lys val lys leu

3782                                                    3812
CTT AAG GAA GTT TTC AGG CAA TTT GTG GAA GCT TGC GGT CAA GCC TTA GCG GTA AAC GAA
leu lys glu val phe arg gln phe val glu ala cys gly gln ala leu ala val asn glu

3842                                                    3872
CGT CTG ATT AAA GAA GAC CAG CTC GAG TAT CAG GAA GAA ATG AAA GCC AAC TAC AGG GAA
arg leu ile lys glu asp gln leu glu tyr gln glu glu met lys ala asn tyr arg glu
```

Insertion of 8 nucleotides found only in Human CLASP-2D with sequence: CTGGGATG

```
3902                                                    3932
ATG GCG AAG GAG CTT TCT GAA ATC ATG CAT GAG CAG ATC TGC CCC CTG GAG GAG AAG ACG
met ala lys glu leu ser glu ile met his glu gln ile cys pro leu glu glu lys thr

3962                                                    3992
AGC GTC TTA CCG AAT TCC CTT CAC ATC TTC AAC GCC ATC AGT GGG ACT CCA ACA AGC ACA
ser val leu pro asn ser leu his ile phe asn ala ile ser gly thr pro thr ser thr

4022                                        |xxxxx PBM xxxxx|
ATG GTT CAC GGG ATG ACC AGC TCG TCT TCG GTC GTG TGA TTA CAT CTC ATG GCC CGT GTG
met val his gly met thr ser ser ser ser val val STP

4082                                                    4112
TGG GGA CTT GCT TTG TCA TTT GCA AAC TCA GGA TGC TTT CCA AAG CCA ATC ACT GGG GAG

4142                                                    4172
ACC GAG CAC AGG GAG GAC CAA GGG GAA GGG GAG AGA AAG GAA ATA AAG AAC AAC GTT ATT

4202                                                    4232
TCT TAA CAG ACT TTC TAT AGG AGT TGT AAG AAG GTG CAC ATA TTT TTT TAA ATC TCA CTG

4262                                                    4292
GCA ATA TTC AAA GTT TTC ATT GTG TCT TAA CAA AGG TGT GGT AGA CAC TCT TGA GCT GGA

4322                                                    4352
CTT AGA TTT TAT TCT TCC TTG CAG AGT AGT GTT AGA ATA GAT GGC CTA CAG AAA AAA AAG

4382                                                    4412
GTT CTG GGA TCT ACA TGG CAG GGA GGG CTG CAC TGA CAT TGA TGC CTG GGG GAC CTT TTG
```

FIG. 2B (cont.)

```
4462                                                4472
CCT CGA CTC GTG CCG GAA ATC TGA TCG TAA TCA GGG TAC AGA ACT TAC TAG TTT TGT CTA

4502                                                4532
GGA GTA TGT TGT ATG ACT AGG ATT TGT GCT ATT ATC TCA TTC AAC AAC ATA GAG CAA GAA

4562                                                4592
TAG TGA GCT AAC TCA CCT AGA CAC TCA ATT AAT CCG CTA CTG GCT TCA AGT CAG AAC TTT

4622                                                4652
GTC ATT AAT CAT CGA CTC CGG GAC GGT CAT ATA TGT ATT ACA TTT CTA CAT TTT TAA TAC

4682                                                4712
TCA CAT CCC CTT ATC CAT TAA GTT TAA TTG TGA TAA ATT TGT GCT GGT CCA GTA TAT GCA

4742                                                4772
ATA CAC TTT AAT GGT TTA TTC TTG TCA TAA AAA TGT GCA ATA TGG AGA TGT ATA CAA GTC

4802
TTT ACT
```

FIG. 2B (cont.)

```
HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   GCATCTGGAAATCTTGACAAAAATGCCAGATTTTCTGCCATCTACAGGCAAGACAGCAAT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   AAGCTATCCAATGATGACATGCTCAAGTTACTTGCAGACTTTCGGAAACCTGAGAAGATG
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   GCTAAGCTCCCAGTGATTTTAGGCAATCTAGACATTACAATTGATAATGTTTCCTCAGAC
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TTCCCTAATTATGTTAATTCATCATACATTCCCACAAAACAATTTGAAACCTGCAGTAAA
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ACTCCCATCACGTTTGAAGTGGAGGAATTTGTGCCCTGCATACCAAAACACACTCAGCCT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------
```

FIG. 3A

```
HC2A              ------------------------------------------------------------
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     TACACCATCTACACCAATCACCTTTACGTTTATCCTAAGTACTTGAAATACGACAGTCAG
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              ------------------------------------------------------------
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AAGTCTTTTGCCAAGGCTAGAAATATTGCGATTTGCATTGAATTCAAAGATTCAGATGAG
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              ------------------------------------------------------------
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GAAGACTCTCAGCCCCTTAAGTGCATTTATGGCAGACCTGGTGGGCCAGTTTTCACAAGA
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              ------------AGTTTTACACCATCACCAAAACCCAGAATTTTATGATGAGATTAAA
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AGCGCCTTTGCTGCAGTTTTACACCATCACCAAAACCCAGAATTTTATGATGAGATTAAA
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              ATAGAGTTGCCCACTCAGCTGCATGAAAAGCACCACCTGTTGCTCACATTCTTCCATGTC
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     ATAGAGTTGCCCACTCAGCTGCATGAAAAGCACCACCTGTTGCTCACATTCTTCCATGTC
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              AGCTGTGACAACTCAAGTAAAGGAAGCACGAAGAAGAGGGATGTCGTTGAAACCCAAGTT
HC2-80            ------------------------------------------------------------
HC2B              ------------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AGCTGTGACAACTCAAGTAAAGGAAGCACGAAGAAGAGGGATGTCGTTGAAACCCAAGTT
HC2E              ------------------------------------------------------------
HC2F              ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A            GGCTACTCCTGGCTTCCCCTCCTGAAAGACGGAAGGGTGGTGACAAGCGAGCAGCACATC
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   GGCTACTCCTGGCTTCCCCTCCTGAAAGACGGAAGGGTGGTGACAAGCGAGCAGCACATC
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CCGGTCTCGGCGAACCTTCCTTCGGGCTATCTTGGCTACCAAGAGCTTGGGATGGGCAGG
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CCGGTCTCGGCGAACCTTCCTTCGGGCTATCTTGGCTACCAGGAGCTTGGGATGGGCAGG
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CATTATGGTCCGGAAATTAAATGGGTAGATGGAGGCAAGCCACTGCTGAAAATTTCCACT
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CATTATGGTCCGGAAATTAAATGGGTAGATGGAGGCAAGCCACTGCTGAAAATTTCCACT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CATCTGGTTTCTACAGTGTATACTCAGGATCAGCATTTACATAATTTTTTCCAGTACTGT
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CATCTGGTTTCTACAGTGTATACTCAGGATCAGCATTTACATAATTTTTTCCAGTACTGT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CAGAAAACCGAATCTGGAGCCCAAGCCTTAGGAAACGAACTTGTAAAGTACCTTAAGAGT
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CAGAAAACCGAATCTGGAGCCCAAGCCTTAGGAAACGAACTTGTAAAGTACCTTAAGAGT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CTGCATGCGATGGAAGGCCACGTGATGATCGCCTTCTTGCCCACTATCCTAAACCAGCTG
HC2-80          ------------------------------------------------------------
HC2B            ------GCGATGGAAGGCCACGTGATGATCGCCTTCTTGCCCACTATCCTAAACCAGCTG
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CTGCATGCGATGGAAGGCCACGTGATGATCGCCTTCTTGCCCACTATCCTAAACCAGCTG
HC2E            ------GCGATGGAAGGCCACGTGATGATCGCCTTCTTGCCCACTATCCTAAACCAGCTG
HC2F            ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A            TTCCGAGTCCTCACCAGAGCCACACAGGAAGAAGTCGCGGTTAACGTGACTCGGGTCATT
HC2-80          ------------------------------------------------------------
HC2B            TTCCGAGTCCTCACCAGAGCCACACAGGAAGAAGTCGCGGTTAACGTGACTCGGGTCATT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TTCCGAGTCCTCACCAGAGCCACACAGGAAGAAGTCGCGGTTAACGTGACTCGGGTCATT
HC2E            TTCCGAGTCCTCACCAGAGCCACACAGGAAGAAGTCGCGGTTAACGTGACTCGGGTCATT
HC2F            ------------------------------------------------------------

HC2A            ATTCATGTGGTTGCCCAGTGCCATGAGGAAGGATTGGAGAGCCACTTGAGGTCATATGTT
HC2-80          ------------------------------------------------------------
HC2B            ATTCATGTGGTTGCCCAGTGCCATGAGGAAGGATTGGAGAGCCACTTGAGGTCATATGTT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ATTCATGTGGTTGCCCAGTGCCATGAGGAAGGATTGGAGAGCCACTTGAGGTCATATGTT
HC2E            ATTCATGTGGTTGCCCAGTGCCATGAGGAAGGATTGGAGAGCCACTTGAGGTCATATGTT
HC2F            ------------------------------------------------------------

HC2A            AAGTACGCGTATAAGGCTGAGCCATATGTTGCCTCTGAATACAAGACAGTGCATGAAGAA
HC2-80          ------------------------------------------------------------
HC2B            AAGTACGCGTATAAGGCTGAGCCATATGTTGCCTCTGAATACAAGACAGTGCATGAAGAA
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   AAGTACGCGTATAAGGCTGAGCCATATGTTGCCTCTGAATACAAGACAGTGCATGAAGAA
HC2E            AAGTACGCGTATAAGGCTGAGCCATATGTTGCCTCTGAATACAAGACAGTGCATGAAGAA
HC2F            ------------------------------------------------------------

HC2A            CTGACCAAATCCATGACCACGATTCTCAAGCCTTCTGCCGATTTCCTCACCAGCAACAAA
HC2-80          ------------------------------------------------------------
HC2B            CTGACCAAATCCATGACCACGATTCTCAAGCCTTCTGCCGATTTCCTCACCAGCAACAAA
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CTGACCAAATCCATGACCACGATTCTCAAGCCTTCTGCCGATTTCCTCACCAGCAACAAA
HC2E            CTGACCAAATCCATGACCACGATTCTCAAGCCTTCTGCCGATTTCCTCACCAGCAACAAA
HC2F            ------------------------------------------------------------

HC2A            CTACTGAGGTACTCATGGTTTTTCTTTGATGTACTGATCAAATCTATGGCTCAGCATTTG
HC2-80          ------------------------------------------------------------
HC2B            CTACTGAGGTACTCATGGTTTTTCTTTGATGTACTGATCAAATCTATGGCTCAGCATTTG
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CTACTGAAGTACTCATGGTTTTTCTTTGATGTACTGATCAAATCTATGGCTCAGCATTTG
HC2E            CTACTGAGGTACTCATGGTTTTTCTTTGATGTACTGATCAAATCTATGGCTCAGCATTTG
HC2F            ------------------------------------------------------------

HC2A            ATAGAGAACTCCAAAGTTAAGTTGCTGCGAAACCAGAGATTTCCTGCATCCTATCATCAT
HC2-80          ------------------------------------------------------------
HC2B            ATAGAGAACTCCAAAGTTAAGTTGCTGCGAAACCAGAGATTTCCTGCATCCTATCATCAT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ATAGAGAACTCCAAAGTTAAGTTGCTGCGAAACCAGAGATTTCCTGCATCCTATCATCAT
HC2E            ATAGAGAACTCCAAAGTTAAGTTGCTGCGAAACCAGAGATTTCCTGCATCCTATCATCAT
HC2F            ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A              GCAGCGGAAACCGTTGTAAATATGCTGATGCCACACATCACTCAGAAGTTTGGAGATAAT
HC2-80            ------------------------------------------------------------
HC2B              GCAGCGGAAACCGTTGTAAATATGCTGATGCCACACATCACTCAGAAGTTTGGAGATAAT
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GCAGTGGAAACCGTTGTAAATATGCTGATGCCACACATCACTCAGAAGTTTCGAGATAAT
HC2E              GCAGCGGAAACCGTTGTAAATATGCTGATGCCACACATCACTCAGAAGTTTGGAGATAAT
HC2F              ------------------------------------------------------------

HC2A              CCAGAGGCATCTAAGAACGCGAATCATAGCCTTGCTGTCTTCATCAAGAGATGTTTCACC
HC2-80            ------------------------------------------------------------
HC2B              CCAGAGGCATCTAAGAACGCGAATCATAGCCTTGCTGTCTTCATCAAGAGATGTTTCACC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     CCAGAGGCATCTAAGAACGCGAATCATAGCCTTGCTGTCTTCATCAAGAGATGTTTCACC
HC2E              CCAGAGGCATCTAAGAACGCGAATCATAGCCTTGCTGTCTTCATCAAGAGATGTTTCACC
HC2F              ------------------------------------------------------------

HC2A              TTCATGGACAGGGGCTTTGTCTTCAAGCAGATCAACAACTACATTAGCTGTTTTGCTCCT
HC2-80            ------------------------------------------------------------
HC2B              TTCATGGACAGGGGCTTTGTCTTCAAGCAGATCAACAACTACATTAGCTGTTTTGCTCCT
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     TTCATGGACAGGGGCTTTGTCTTCAAGCAGATCAACAACTACATTAGCTGTTTTGCTCCT
HC2E              TTCATGGACAGGGGCTTTGTCTTCAAGCAGATCAACAACTACATTAGCTGTTTTGCTCCT
HC2F              ------------------------------------------------------------

HC2A              GGAGACCCAAAGACCCTCTTTGAATACAAGTTTGAATTTCTCCGTGTAGTGTGCAACCAT
HC2-80            ------------------------------------------------------------
HC2B              GGAGACCCAAAGACCCTCTTTGAATACAAGTTTGAATTTCTCCGTGTAGTGTGCAACCAT
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GGAGACCCAAAGACCCTCTTTGAATACAAGTTTGAATTTCTCCGTGTAGTGTGCAACCAT
HC2E              GGAGACCCAAAGACCCTCTTTGAATACAAGTTTGAATTTCTCCGTGTAGTGTGCAACCAT
HC2F              ------------------------------------------------------------

HC2A              GAACATTATATTCCGTTGAACTTACCAATGCCATTTGGAAAAGGCAGGATTCAAAGATAC
HC2-80            ------------------------------------------------------------
HC2B              GAACATTATATTCCGTTGAACTTACCAATGCCATTTGGAAAAGGCAGGATTCAAAGATAC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GAACATTATATTCCGTTGAACTTACCAATGCCATTTGGAAAAGGCAGGATTCAAAGATAC
HC2E              GAACATTATATTCCGTTGAACTTACCAATGCCATTTGGAAAAGGCAGGATTCAAAGATAC
HC2F              ------------------------------------------------------------

HC2A              CAAGACCTCCAGCTTGACTACTCATTAACAGATGAGTTCTGCAGAAACCACTTCTTGGTG
HC2-80            -------TCCAGCTTGACTACTCATTAACAGATGAGTTCTGCAGAAACCACTTCTTGGTG
HC2B              CAAGACCTCCAGCTTGACTACTCATTAACAGATGAGTTCTGCAGAAACCACTTCTTGGTG
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     CAAGACCTCCAGCTTGACTACTCATTAACAGATGAGTTCTGCAGAAACCACTTCTTGGTG
HC2E              CAAGACCTCCAGCTTGACTACTCATTAACAGATGAGTTCTGCAGAAACCACTTCTTGGTG
HC2F              ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A              GGACTGTTACTGAGGGAGGTGGGGACAGCCCTCCAGGAGTTCCGGGAGGTCCGTCTGATC
HC2-80            GGACTGTTACTGAGGGAGGTGGGGACAGCCCTCCAGGAGTTCCGGGAGGTCCGTCTGATC
HC2B              GGACTGTTACTGAGGGAGGTGGGGACAGCCCTCCAGGAGTTCCGGGAGGTCCGTCTGATC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GGACTGTTACTGAGGGAGGTGGGGACAGCCCTCCAGGAGTTCCGGGAGGTCCGTCTGATC
HC2E              GGACTGTTACTGAGGGAGGTGGGGACAGCCCTCCAGGAGTTCCGGGAGGTCCGTCTGATC
HC2F              ------------------------------------------------------------

HC2A              GCCATCAGTGTGCTCAAGAACCTGCTGATAAAGCATTCTTTTGATGACAGATATGCTTCA
HC2-80            GCCATCAGTGTGCTCAAGAACCTGCTGATAAAGCATTCTTTTGATGACAGATATGCTTCA
HC2B              GCCATCAGTGTGCTCAAGAACCTGCTGATAAAGCATTCTTTTGATGACAGATATGCTTCA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GCCATCAGTGTGCTCAAGAACCTGCTGATAAAGCATTCTTTTGATGACAGATATGCTTCA
HC2E              GCCATCAGTGTGCTCAAGAACCTGCTGATAAAGCATTCTTTTGATGACAGATATGCTTCA
HC2F              ------------------------------------------------------------

HC2A              AGGAGCCATCAGGCAAGGATAGCCACCCTCTACCTGCCTCTGTTTGGTCTGCTGATTGAA
HC2-80            AGGAGCCATCAGGCAAGGATAGCCACCCTCTACCTGCCTCTGTTTGGTCTGCTGATTGAA
HC2B              AGGAGCCATCAGGCAAGGATAGCCACCCTCTACCTGCCTCTGTTTGGTCTGCTGATTGAA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AGGAGCCATCAGGCAAGGATAGCCACCCTCTACCTGCCTCTGTTTGGTCTGCTGATTGAA
HC2E              AGGAGCCATCAGGCAAGGATAGCCACCCTCTACCTGCCTCTGTTTGGTCTGCTGATTGAA
HC2F              ------------------------------------------------------------

HC2A              AACGTCCAGCGGATCAATGTGAGGGATGTGTCACCCTTCCCTGTGAACGCGGGCATGACC
HC2-80            AACGTCCAGCGGATCAATGTGAGGGATGTGTCACCCTTCCCTGTGAACGCGGGCATGACC
HC2B              AACGTCCAGCGGATCAATGTGAGGGATGTGTCACCCTTCCCTGTGAACGCGGGCATGACC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AACGTCCAGCGGATCAATGTGAGGGATGTGTCACCCTTCCCTGTGAACGCGGGCATGACT
HC2E              AACGTCCAGCGGATCAATGTGAGGGATGTGTCACCCTTCCCTGTGAACGCGGGCATGACC
HC2F              ------------------------------------------------------------

HC2A              GTGAAGGATGAATCCCTGGCTCTACCAGCTGTGAATCCGCTGGTGACGCCGCAGAAGGGA
HC2-80            GTGAAGGATGAATCCCTGGCTCTACCAGCTGTGAATCCGCTGGTGACGCCGCAGAAGGGA
HC2B              GTGAAGGATGAATCCCTGGCTCTACCAGCTGTGAATCCGCTGGTGACGCCGCAGAAGGGA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GTGAAGGATGAATCCCTGGCTCTACCAGCTGTGAATCCGCTGGTGACGCCGCAGAAGGGA
HC2E              GTGAAGGATGAATCCCTGGCTCTACCAGCTGTGAATCCGCTGGTGACGCCGCAGAAGGGA
HC2F              ------------------------------------------------------------

HC2A              AGCACCCTGGACAACAGCCTGCACAAGGACCTGCTGGGCGCCATCTCCGGCATTGCTTCT
HC2-80            AGCACCCTGGACAACAGCCTGCACAAGGACCTGCTGGGCGCCATCTCCGGCATTGCTTCT
HC2B              AGCACCCTGGACAACAGCCTGCACAAGGACCTGCTGGGCGCCATCTCCGGCATTGCTTCT
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     AGCACCCTGGACAACAGCCTGCACAAGGACCTGCTGGGCGCCATCTCCGGCATTGCTTCT
HC2E              AGCACCCTGGACAACAGCCTGCACAAGGACCTGCTGGGCGCCATCTCCGGCATTGCTTCT
HC2F              ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A                CCATATACAACCTCAACTCCAAACATCAACAGTGTGAGAAATGCTGATTCGAGAGGATCT
HC2-80              CCATATACAACCTCAACTCCAAACATCAACAGTGTGAGAAATGCTGATTCGAGAGGATCT
HC2B                CCATATACAACCTCAACTCCAAACATCAACAGTGTGAGAAATGCTGATTCGAGAGGATCT
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       CCATATACAACCTCAACTCCAAACATCAACAGTGTGAGAAATGCTGATTCGAGAGGATCT
HC2E                CCATATACAACCTCAACTCCAAACATCAACAGTGTGAGAAATGCTGATTCGAGAGGATCT
HC2F                ----------------------------------------GCTGATTCGAGAGGATCT

HC2A                CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC
HC2-80              CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC
HC2B                CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC
HC2E                CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC
HC2F                CTCATAAGCACAGATTCGGGTAACAGCCTTCCAGAAAGGAATAGTGAGAAGAGCAATTCC

HC2A                CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT
HC2-80              CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT
HC2B                CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT
HC2E                CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT
HC2F                CTGGATAAGCACCAACAAAGTAGCACATTGGGAAATTCCGTGGTTCGCTGTGATAAACTT

HC2A                GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT
HC2-80              GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT
HC2B                GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT
HC2E                GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT
HC2F                GACCAGTCTGAGATTAAGAGCCTACTGATGTGTTTCCTCTACATCTTAAAGAGCATGTCT

HC2A                GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT
HC2-80              GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT
HC2B                GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT
HC2E                GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT
HC2F                GATGATGCTTTGTTTACATATTGGAACAAGGCTTCAACATCTGAACTTATGGATTTTTTT

HC2A                ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAGG
HC2-80              ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAGG
HC2B                ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAGG
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAG-
HC2E                ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAGG
HC2F                ACAATATCTGAAGTCTGCCTGCACCAGTTCCAGTACATGGGGAAGCGATACATAGCCAG-
```

FIG. 3A (cont.)

```
HC2A            AACCAGGAGGGGTTGGGACCCATAGTTCATGATCGAAAGTCTCAGACATTGCCTGTTTCC
HC2-80          AACCAGGAGGGGTTGGGACCCATAGTTCATGATCGAAAGTCTCAGACATTGCCTGTTTCC
HC2B            AACCAGGAGGGGTTGGGACCCATAGTTCATGATCGAAAGTCTCAGACATTGCCTGTTTCC
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ----------------------------------AA----------------------
HC2E            AACCAGGAGGGGTTGGGACCCATAGTTCATGATCGAAAGTCTCAGACATTGCCTGTTTCC
HC2F            ----------TGTGA----------------GAAAG------ATATCAAGTGT----

HC2A            CGTAACAGAACAGGAATGATGCATGCCAGATTGCAGCAGCTGGGCAGCCTGGATAACTCT
HC2-80          CGTAACAGAACAGGAATGATGCATGCCAGATTGCAGCAGCTGGGCAGCCTGGATAACTCT
HC2B            CGTAACAGAACAGGAATGATGCATGCCAGATTGCAGCAGCTGGGCAGCCTGGATAACTCT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ----------CAGGAATGATGCATGCCAGATTGCAGCAGCTGGGCAGCCTGGATAACTCT
HC2E            CGTAACAGAACAGGAATGATGCATGCCAGATTGCAGCAGCTGGGCAGCCTGGATAACTCT
HC2F            ----------------------------------------GCTTGGAA----------

HC2A            CTCACTTTTAACCACAGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT
HC2-80          CTCACTTTTAACCACAGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT
HC2B            CTCACTTTTAACCACAGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CTCACTTTTAACCACAGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT
HC2E            CTCACTTTTAACCACAGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT
HC2F            -TTTCTGTAGACAATGGCTATGGCCACTCGGACGCAGATGTTCTGCACCAGTCATTACTT

HC2A            GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA
HC2-80          GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA
HC2B            GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA
HC2E            GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA
HC2F            GAAGCCAACATTGCTACTGAGGTTTGCCTGACAGCTCTGGACACGCTTTCTCTATTTACA

HC2A            TTGGCGTTTAAGAACCAGCTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAGTT
HC2-80          TTGGCGTTTAAGAACCAGCTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAGTT
HC2B            TTGGCGTTTAAG------CTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAGTT
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TTGGCGTTTAAGAACCAGCTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAGTT
HC2E            TTGGCGTTTAAGAACCAGCTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAGTT
HC2F            TTGGCGTTTAAGAACCAGCTCCTGGCCGACCATGGACATAATCCTCTCATGAAAAAAAAA

HC2A            TTTGATGTCTACCTGTGTTTTCTTCAAAAACATCAGTCTGAAACGGCTTTAAAAAATGTC
HC2-80          TTTGATGTCTACCTGTGTTTTCTTCAAAAACATCAGTCTGAAACGGCTTTAAAAAATGTC
HC2B            TTTGATGTCTACCTGTGTTTTCTTCAAAAACATCAGTCTGAAACGGCTTTAAAAAATGTC
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TTTGATGTCTACCTGTGTTTTCTTCAAAAACATCAGTCTGAAACGGCTTTAAAAAATGTC
HC2E            TTTGATGTCTACCTGTGTTTTCTTCAAAAACATCAGTCTGAAACGGCTTTAAAAAATGTC
HC2F            A-----------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A              TTCACTGCCTTAAGGTCCTTAATTTATAAGTTTCCCTCAACATTCTATGAAGGGAGAGCG
HC2-80            TTCACTGCCTTAAGGTCCTTAATTTATAAGTTTCCCTCAACATTCTATGAAGGGAGAGCG
HC2B              TTCACTGCCTTAAGGTCCTTAATTTATAAGTTTCCCTCAACATTCTATGAAGGGAGAGCG
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     TTCACTGCCTTAAGGTCCTTAATTTATAAGTTTCCCTCAACATTCTATGAAGGGAGAGCG
HC2E              TTCACTGCCTTAAGGTCCTTAATTTATAAGTTTCCCTCAACATTCTATGAAGGGAGAGCG
HC2F              ------------------------------------------------------------

HC2A              GACATGTGTGCGGCTCTGTGTTACGAGATTCTCAAGTGCTGTAACTCCAAGCTGAGCTCC
HC2-80            GACATGTGTGCGGCTCTGTGTTACGAGATTCTCAAGTGCTGTAACTCCAAGCTGAGCTCC
HC2B              GACATGTGTGCGGCTCTGTGTTACGAGATTCTCAAGTGCTGTAACTCCAAGCTGAGCTCC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GACATGTGTGCGGCTCTGTGTTACGAGATTCTCAAGTGCTGTAACTCCAAGCTGAGCTCC
HC2E              GACATGTGTGCGGCTCTGTGTTACGAGATTCTCAAGTGCTGTAACTCCAAGCTGAGCTCC
HC2F              ------------------------------------------------------------

HC2A              ATCAGGACGGAGGCCTCCCAGCTGCTCTACTTCCTGATGAGGAACAACTTTGATTACACT
HC2-80            ATCAGGACGGAGGCCTCCCAGCTGCTCTACTTCCTGATGAGGAACAACTTTGATTACACT
HC2B              ATCAGGACGGAGGCCTCCCAGCTGCTCTACTTCCTGATGAGGAACAACTTTGATTACACT
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     ATCAGGACGGAGGCCTCCCAGCTGCTCTACTTCCTGATGAGGAACAACTTTGATTACACT
HC2E              ATCAGGACGGAGGCCTCCCAGCTGCTCTACTTCCTGATGAGGAACAACTTTGATTACACT
HC2F              ------------------------------------------------------------

HC2A              GGAAAGAAGTCCTTTGTCCGGACACATTTGCAAGTCATCATATCTGTCAGCCAGCTGATA
HC2-80            GGAAAGAAGTCCTTTGTCCGGACACATTTGCAAGTCATCATATCTGTCAGCCAGCTGATA
HC2B              GGAAAGAAGTCCTTTGTCCGGACACATTTGCAAGTCATCATATCTGTCAGCCAGCTGATA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GGAAAGAAGTCCTTTGTCCGGACACATTTGCAAGTCATCATATCTGTCAGCCAGCTGATA
HC2E              GGAAAGAAGTCCTTTGTCCGGACACATTTGCAAGTCATCATATCTGTCAGCCAGCTGATA
HC2F              ------------------------------------------------------------

HC2A              GCAGACGTTGTTGGCATTGGGGAAACCAGATTCCAGCAGTCCCTGTCCATCATCAACAAC
HC2-80            GCAGACGTTGTTGGCATTGGGGAAACCAGATTCCAGCAGTCCCTGTCCATCATCAACAAC
HC2B              GCAGACGTTGTTGGCATTGGGGAAACCAGATTCCAGCAGTCCCTGTCCATCATCAACAAC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GCAGACGTTGTTGGCATTGGGGGAACCAGATTCCAGCAGTCCCTGTCCATCATCAACAAC
HC2E              GCAGACGTTGTTGGCATTGGGGAAACCAGATTCCAGCAGTCCCTGTCCATCATCAACAAC
HC2F              ------------------------------------------------------------

HC2A              TGTGCCAACAGTGACCGGCTTATTAAGCACACCAGCTTCTCCTCTGATGTGAAGGACTTA
HC2-80            TGTGCCAACAGTGACCGGCTTATTAAGCACACCAGCTTCTCCTCTGATGTGAAGGACTTA
HC2B              TGTGCCAACAGTGACCGGCTTATTAAGCACACCAGCTTCTCCTCTGATGTGAAGGACTTA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     TGTGCCAACAGTGACCGGCTTATTAAGCACACCAGCTTCTCCTCTGATGTGAAGGACTTA
HC2E              TGTGCCAACAGTGACCGGCTTATTAAGCACACCAGCTTCTCCTCTGATGTGAAGGACTTA
HC2F              ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A              ACCAAAAGGATACGCACGGTGCTAATGGCCACCGCCCAGATGAAGGAGCATGAGAACGAC
HC2-80            ACCAAAAGGATACGCACGGTGCTAATGGCCACCGCCCAGATGAAGGAGCATGAGAACGAC
HC2B              ACCAAAAGGATACGCACGGTGCTAATGGCCACCGCCCAGATGAAGGAGCATGAGAACGAC
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     ACCAAAAGGATACGCACGGTGCTAATGGCCACCGCCCAGATGAAGGAGCATGAGAACGAC
HC2E              ACCAAAAGGATACGCACGGTGCTAATGGCCACCGCCCAGATGAAGGAGCATGAGAACGAC
HC2F              ------------------------------------------------------------

HC2A              CCAGAGATGCTGGTGGACCTCCAGTACAGCCTGGCCAAATCCTATGCCAGCACGCCCGAG
HC2-80            CCAGAGATGCTGGTGGACCTCCAGTACAGCCTGGCCAAATCCTATGCCAGCACGCCCGAG
HC2B              CCAGAGATGCTGGTGGACCTCCAGTACAGCCTGGCCAAATCCTATGCCAGCACGCCCGAG
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     CCAGAGATGCTGGTGGACCTCCAGTACAGCCTGGCCAAATCCTATGCCAGCACGCCCGAG
HC2E              CCAGAGATGCTGGTGGACCTCCAGTACAGCCTGGCCAAATCCTATGCCAGCACGCCCGAG
HC2F              ------------------------------------------------------------

HC2A              CTCAGGAAGACGTGGCTCGACAGCATGGCCAGGATCCATGTCAAAAATGGCGATCTCTCA
HC2-80            CTCAGGAAGACGTGGCTCGACAGCATGGCCAGGATCCATGTCAAAAATGGCGATCTCTCA
HC2B              CTCAGGAAGACGTGGCTCGACAGCATGGCCAGGATCCATGTCAAAAATGGCGATCTCTCA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     CTCAGGAAGACGTGGCTCGACAGCATGGCCAGGATCCATGTCAAAAATGGCGATCTCTCA
HC2E              CTCAGGAAGACGTGGCTCGACAGCATGGCCAGGATCCATGTCAAAAATGGCGATCTCTCA
HC2F              ------------------------------------------------------------

HC2A              GAGGCAGCAATGTGCTATGTCCACGTAACAGCCCTAGTGGCAGAATATCTCACACGGAAA
HC2-80            GAGGCAGCAATGTGCTATGTCCACGTAACAGCCCTAGTGGCAGAATATCTCACACGGAAA
HC2B              GAGGCAGCAATGTGCTATGTCCACGTAACAGCCCTAGTGGCAGAATATCTCACACGGAAA
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GAGGCAGCAATGTGCTATGTCCACGTAACAGCCCTAGTGGCAGAATATCTCACACGGAAA
HC2E              GAGGCAGCAATGTGCTATGTCCACGTAACAGCCCTAGTGGCAGAATATCTCACACGGAAA
HC2F              ------------------------------------------------------------

HC2A              G-----------------------------------------------------------
HC2-80            G-----------------------------------------------------------
HC2B              G-----------------------------------------------------------
HC2C              ------------------------------------------------------------
HC2D-KIAA1058     GAAGCAGTCCAGTGGGAGCCGCCCCTTCTCCCCCACAGCCATAGCGCCTGCCTGAGGAGG
HC2E              G-----------------------------------------------------------
HC2F              ------------------------------------------------------------

HC2A              ----------GCGTGTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2-80            ----------GCGTGTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2B              ----------GCGTGTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2C              --------------GTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2D-KIAA1058     AGCCGGGGAGGCGTGTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2E              ----------GCGTGTTTAGACAAGGATGCACCGCCTTCAGGGTCATTACCCCAAACATC
HC2F              ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A              GACGAGGAGGCCTCCATGATGGAAGACGTGGGGATGCAGGATGTCCATTTCAACGAGGAT
HC2-80            GACGAGGAGGCCTCCATGATGGAAGACGTGGGGATGCAGGATGTCCATTTCAACGAGGAT
HC2B              GACGAGGAGGCCTCCATGATGGAAGACGTGGGGATGCAGGATGTCCATTTCAACGAGGAT
HC2C              GACGAGGAGGCCTCCATGATGGAAGACGTGGGGATGCAGGATGTCCATTTCAACGAGGAT
HC2D-KIAA1058     GACGAGGAGGCCTCCATGATGGAAGACGTGGGGATGCAGGATGTCCATTTCAACGAGGAT
HC2E              GACGAGGAGGCCTCCATGATGGAAGACGTGGGGA--------------------------
HC2F              --------------------------------------------------------

HC2A              GTGCTGATGGAGCTCCTTGAGCAGTGCGCAGATGGACTCTGGAAAGCCGAGCGCTACGAG
HC2-80            GTGCTGATGGAGCTCCTTGAGCAGTGCGCAGATGGACTCTGGAAAGCCGAGCGCTACGAG
HC2B              GTGCTGATGGAGCTCCTTGAGCAGTGCGCAGATGGACTCTGGAAAGCCGAGCGCTACGAG
HC2C              GTGCTGATGGAGCTCCTTGAGCAGTGCGCAGATGGACTCTGGAAAGCCGAGCGCTACGAG
HC2D-KIAA1058     GTGCTGATGGAGCTCCTTGAGCAGTGCGCAGATGGACTCTGGAAAGCCGAGCGCTACGAG
HC2E              ------------------------------------------AAGCCGAGCGCTACGAG
HC2F              --------------------------------------------------------

HC2A              CTCATCGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTT-----
HC2-80            CTCATCGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTT-----
HC2B              CTCATCGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTTTGAG
HC2C              CTCATCGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTTTGAG
HC2D-KIAA1058     CTCATTGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTTTGAG
HC2E              CTCATCGCCGACATCTACAAACTTATCATCCCCATTTATGAGAAGCGGAGGGATTTTGAG
HC2F              --------------------------------------------------------

HC2A              --------------------------------------------------------
HC2-80            --------------------------------------------------------
HC2B              AGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGACCGAGGTCATG
HC2C              AGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGACCGAGGTCATG
HC2D-KIAA1058     AGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGACCGAGGTCATG
HC2E              AGGCTGGCCCATCTGTATGACACGCTGCACCGGGCCTACAGCAAAGTGACCGAGGTCATG
HC2F              --------------------------------------------------------

HC2A              --------------------------------------------------------
HC2-80            --------------------------------------------------------
HC2B              CACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTCTTCGGGCAGG-----
HC2C              CACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTCTTCGGGCAGG-----
HC2D-KIAA1058     CACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTCTTCGGGCAGGCAGCG
HC2E              CACTCGGGCCGCAGGCTTCTGGGGACCTACTTCCGGGTAGCCTTCTTCGGGCAGG-----
HC2F              --------------------------------------------------------

HC2A              ------------------------------------------CTTTGAAGATGAAGATGGA
HC2-80            ------------------------------------------CTTTGAAGATGAAGATGGA
HC2B              ----------------------------------------GATTCTTTGAAGATGAAGATGGA
HC2C              ----------------------------------------GATTCTTTGAAGATGAAGATGGA
HC2D-KIAA1058     CAATACCAGTTTACAGACAGTGAAACAGATGTGGAGGGATTCTTTGAAGATGAAGATGGA
HC2E              ----------------------------------------GATTCTTTGAAGATGAAGATGGA
HC2F              --------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A                AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2-80              AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2B                AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2C                AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2D-KIAA1058       AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2E                AAGGAGTATATTTACAAGGAACCCAAACTCACACCGCTGTCGGAAATTTCTCAGAGACTC
HC2F                ------------------------------------------------------------

HC2A                CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGC
HC2-80              CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGC
HC2B                CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGACACAGGATTCTGGC
HC2C                CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGC
HC2D-KIAA1058       CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGC
HC2E                CTTAAACTGTACTCGGATAAATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGC
HC2F                ------------------------------------------------------------

HC2A                AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCATACATCCAGGTGACTCACGTCATC
HC2-80              AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCATACATCCAGGTGACTCACGTCATC
HC2B                AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCATACATCCAGGTGACTCACGTCATC
HC2C                AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCATACATCCAGGTGACTCACGTCATC
HC2D-KIAA1058       AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCCTACATCCAGGTGACTCACGTCATC
HC2E                AAGGTCAACCCTAAGGATCTGGATTCTAAGTATGCATACATCCAGGTGACTCACGTCATC
HC2F                ------------------------------------------------------------

HC2A                CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2-80              CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2B                CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2C                CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2D-KIAA1058       CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2E                CCCTTCTTTGACGAAAAAGAGTTGCAAGAAAGGAAAACAGAGTTTGAGAGATCCCACAAC
HC2F                ------------------------------------------------------------

HC2A                ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2-80              ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2B                ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2C                ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2D-KIAA1058       ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2E                ATCCGCCGCTTCATGTTTGAGATGCCATTTACGCAGACCGGGAAGAGGCAGGGCGGGGTG
HC2F                ------------------------------------------------------------

HC2A                GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2-80              GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2B                GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2C                GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2D-KIAA1058       GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2E                GAAGAGCAGTGCAAACGGCGCACCATCCTGACAGCCATACACTGCTTCCCTTATGTGAAG
HC2F                ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A            AAGCGCATCCCTGTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGTGGCCATT
HC2-80          AAGCGCATCCCTGTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGTGGCCATT
HC2B            AAGCGCATCCCTGTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGTGGCCATT
HC2C            AAGCGCATCCCTTTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGT--CCATT
HC2D-KIAA1058   AAGCGCATCCCTGTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGTGGCCATT
HC2E            AAGCGCATCCCTGTCATGTACCAGCACCACACTGACCTGAACCCCATCGAGGTGGCCATT
HC2F            ------------------------------------------------------------

HC2A            GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2-80          GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2B            GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2C            GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2D-KIAA1058   GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2E            GACGAGATGAGTAAGAAGGTGGCGGAGCTCCGGCAGCTGTGCTCCTCGGCCGAGGTGGAC
HC2F            ------------------------------------------------------------

HC2A            ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2-80          ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2B            ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2C            ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2D-KIAA1058   ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2E            ATGATCAAACTGCAGCTCAAACTCCAGGGCAGCGTGAGTGTTCAGGTCAATGCTGGCCCA
HC2F            ------------------------------------------------------------

HC2A            CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2-80          CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2B            CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2C            CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2D-KIAA1058   CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2E            CTAGCATATGCGCGAGCTTTCTTAGATGATACAAACACAAAGCGATATCCTGACAATAAA
HC2F            ------------------------------------------------------------

HC2A            GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2-80          GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2B            GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2C            GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2D-KIAA1058   GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2E            GTGAAGCTGCTTAAGGAAGTTTTCAGGCAATTTGTGGAAGCTTGCGGTCAAGCCTTAGCG
HC2F            ------------------------------------------------------------

HC2A            GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2-80          GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2B            GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2C            GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2D-KIAA1058   GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2E            GTAAACGAACGTCTGATTAAAGAAGACCAGCTCGAGTATCAGGAAGAAATGAAAGCCAAC
HC2F            ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A                TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAG--------ATCTGCC
HC2-80              TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAG--------ATCTGCC
HC2B                TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAG--------ATCTGCC
HC2C                TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAG--------ATCTGCC
HC2D-KIAA1058       TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAGCTGGGATGATCTGCC
HC2E                TACAGGGAAATGGCGAAGGAGCTTTCTGAAATCATGCATGAGCAG--------ATCTGCC
HC2F                ------------------------------------------------------------

HC2A                CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2-80              CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2B                CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2C                CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2D-KIAA1058       CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2E                CCCTGGAGGAGAAGACGAGCGTCTTACCGAATTCCCTTCACATCTTCAACGCCATCAGTG
HC2F                ------------------------------------------------------------

HC2A                GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGATTAC
HC2-80              GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGATTAC
HC2B                GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGA----
HC2C                GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGA----
HC2D-KIAA1058       GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGATTAC
HC2E                GGACTCCAACAAGCACAATGGTTCACGGGATGACCAGCTCGTCTTCGGTCGTGTGA----
HC2F                ------------------------------------------------------------

HC2A                ATCTCATGGCCCGTGTGTGGGGACTTGCTTTGTCATTTGCAAACTCAGGATGCTTTCCAA
HC2-80              ATCTCATGGCCCGTGTGTGGGGACTTGCTTTGTCATTTGCAAACTCAGGATGCTTTCCAA
HC2B                ------------------------------------------------------------
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       ATCTCATGGCCCGTGTGTGGGGACTTGCTTTGTCATTTGCAAACTCAGGATGCTTTCCAA
HC2E                ------------------------------------------------------------
HC2F                ------------------------------------------------------------

HC2A                AGCCAATCACTGGGGAGACCGAGCACAGGGAGGACCAAGGGGAAGGGGAGAGAAAGGAAA
HC2-80              AGCCAATCACTGGGGAGACCGAGCACAGGGAGGACCAAGGGGAAGGGGAGAGAAAGGAAA
HC2B                ------------------------------------------------------------
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       AGCCAATCACTGGGGAGACCGAGCACAGGGAGGACCA-GGGGAAGGGGAGAGAAAGGAAA
HC2E                ------------------------------------------------------------
HC2F                ------------------------------------------------------------

HC2A                TAAAGAACAACGTTATTTCTTAACAGACTTTCTATAGGAGTTGTAAGAAGGTGCACATAT
HC2-80              TAAAGAACAACGTTATTTCTTAACAGACTTTCTATAGGAGTTGTAAGAAGGTGCACATAT
HC2B                ------------------------------------------------------------
HC2C                ------------------------------------------------------------
HC2D-KIAA1058       TAAAGAACAACGTTATTTCTTAACAGACTTTCTATAGGAGTTGTAAGAAGGTGCACATAT
HC2E                ------------------------------------------------------------
HC2F                ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A          TTTTTTAAATCTCACTGGCAATATTCAAAGTTTTCATTGTGTCTTAACAAAGGTGTGGTA
HC2-80        TTTTTTAAATCTCACTGGCAATATTCAAAGTTTTCATTGTGTCTTAACAAAGGTGTGGTA
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 TTTTTTAAATCTCACTGGCAATATTCAAAGTTTTCATTGTGTCTTAACAAAGGTGTGGTA
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------

HC2A          GACACTCTTGAGCTGGACTTAGATTTTATTCTTCCTTGCAGAGTAGTGTTAGAATAGATG
HC2-80        GACACTCTTGAGCTGGACTTAGATTTTATTCTTCCTTGCAGAGTAGTGTTAGAATAGATG
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 GACACTCTTGAGCTGGACTTAGATTTTATTCTTCCTTGCAGAGTAGTGTTAGAATAGATG
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------

HC2A          GCCTACAGAAAAAAAAGGTTCTGGGATCTACATGGCAGGGAGGGCTGCACTGACATTGAT
HC2-80        GCCTACAGAAAAAAAAGGTTCTGGGATCTACATGGCAGGGAGGGCTGCACTGACATTGAT
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 GCCTACAGAAAAAAAAGGTTCTGGGATCTACATGGCAGGGAGGGCTGCACTGACATTGAT
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------

HC2A          GCCTGGGGGACCTTTTGCCTCGACTCGTGCCGGAAATCTGATCGTAATCAGGGTACAGAA
HC2-80        GCCTGGGGGACCTTTTGCCTCGACTCGTGCCGGAAATCTGATCGTAATCAGGGTACAGAA
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 GCCTGGGGGACCTTTTGCCTCGAGGCTGAGCTGGAAAATCTTGAAAATATTTTTT----T
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------

HC2A          CTTACTAGTTTTGTCTAGGAGTATGTTGTATGACTAGGATTTGTGCTATTATCTCATTCA
HC2-80        CTTACTAGTTTTGTCTAGGAGTATGTTGTATGACTAGGATTTGTGCTATTATCTCATTCA
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 TTTCCTGTGGCACATTCAGGTTGAATACAAGAACTATTTTTGTGACTAGTTTTTGATGAC
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------

HC2A          ACAACATAGAGCAAGAATAGTGAGCTAACTGAGCTAGACACTCAATTAATCCGCTACTGG
HC2-80        ACAACATAGAGCAAGAATAGTGAGCTAACTGAGCTAGACACTCAATTAATCCGCTACTGG
HC2B          ------------------------------------------------------------
HC2C          ------------------------------------------------------------
HC2D-KIAA1058 CTAAGGGAACTGACCATTGTAATTTTTGTACCAGTGAACCAGGAGATTTAGTGCTTTTAT
HC2E          ------------------------------------------------------------
HC2F          ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A            CTTCAAGTCAGAACTTTGTCATTAATCATCGACTCCGGGACGGTCATATATGTATTACAT
HC2-80          CTTCAAGTCAGAACTTTGTCATTAATCATCGACTCCGGGACGGTCATATATGTATTACAT
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ATTCATTTCCTTGCATTTAAGAAAATATGAAAGCTTAAGGAATTATGTGAGCTTAAAACT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            TTCTACATTTTTAATACTCACATGGGCTTATGCATTAAGTTTAATTGTGATAAATTTGTG
HC2-80          TTCTACATTTTTAATACTCACATGGGCTTATGCATTAAGTTTAATTGTGATAAATTTGTG
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   AGTCAAGCAGTTTAGAACCAAAGGCCTATATTAATAACCGCAACTATGCTGAAAAGTACA
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            CTGGTCCAGTATATGCAATACACTTTAATGGTTTATTCTTGTCATAAAAATGTGCAATAT
HC2-80          CTGGTCCAGTATATGCAATACACTTTAATGGTTTATTCTTGTCATAAAAATGTGCAATAT
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   AAGTAGTACAGTATATTGTTATGTACATATCATTGTTAATACAGTCCTGGCATTCTGTAC
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            GGAGATGTATACAAGTCTTTACT-----------------------------------
HC2-80          GGAGATGTATACAAGTCTTTACT-----------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   ATATATGTATTACATTTCTACATTTTTAATACTCACATGGGCTTATGCATTAAGTTTAAT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TGTGATAAATTTGTGCTGTTCCAGTATATGCAATACACTTTAATGTTTTATTCTTGTACA
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TAAAAATGTGCAATATGGAGATGTATACAGTCTTTACTATATTAGGTTTATAAACAGTTT
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TAAGAATTTCATCCTTTTGCCAAAATGGTGGAGTATGTAATTGGTAAATCATAAATCCTG
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TGGTGAATGGTGGTGTACTTTAAAGCTGTCACCATGTTATATTTTCTTTTAAGACATTAA
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   TTTAGTAATTTTATATTTGGGAAAATAAAGGTTTTTAATTTTATTTAACTGGAATCACTG
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------

HC2A            ------------------------------------------------------------
HC2-80          ------------------------------------------------------------
HC2B            ------------------------------------------------------------
HC2C            ------------------------------------------------------------
HC2D-KIAA1058   CCCTGCTGTAATTAAACATTCTGTACCACATCTGTATTAAAAAGACATTGCTGACC
HC2E            ------------------------------------------------------------
HC2F            ------------------------------------------------------------
```

FIG. 3A (cont.)

```
HC2A    ----------------------------------------------------------
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------------------------------------------
HC2C    ----------------------------------------------------------
HC2D    ASGNLDKNARFSAIYRQDSNKLSNDDMLKLLADFRKPEKMAKLPVILGNLDITIDNVSSD
HC2E    ----------------------------------------------------------
HC2F    ----------------------------------------------------------

HC2A    ----------------------------------------------------------
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------------------------------------------
HC2C    ----------------------------------------------------------
HC2D    FPNYVNSSYIPTKQFETCSKTPITFEVEEFVPCIPKHTQPYTIYTNHLYVYPKYLKYDSQ
HC2E    ----------------------------------------------------------
HC2F    ----------------------------------------------------------

HC2A    ----------------------------------------VLHHHQNPEFYDEIK
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------------------------------------------
HC2C    ----------------------------------------------------------
HC2D    KSFAKARNIAICIEFKDSDEEDSQPLKCIYGRPGGPVFTRSAFAAVLHHHQNPEFYDEIK
HC2E    ----------------------------------------------------------
HC2F    ----------------------------------------------------------

HC2A    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------------------------------------------
HC2C    ----------------------------------------------------------
HC2D    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
HC2E    ----------------------------------------------------------
HC2F    ----------------------------------------------------------

HC2A    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------------------------------------------
HC2C    ----------------------------------------------------------
HC2D    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
HC2E    ----------------------------------------------------------
HC2F    ----------------------------------------------------------

HC2A    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLTRATQEEVAVNVTRVI
HC2A-80 ----------------------------------------------------------
HC2B    ----------------------AMEGHVMIAFLPTILNQLFRVLTRATQEEVAVNVTRVI
HC2C    ----------------------------------------------------------
HC2D    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLTRATQEEVAVNVTRVI
HC2E    ----------------------AMEGHVMIAFLPTILNQLFRVLTRATQEEVAVNVTRVI
HC2F    ----------------------------------------------------------
```

Fig. 3B

```
HC2A      IHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSNK
HC2A-80   ------------------------------------------------------------
HC2B      IHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSNK
HC2C      ------------------------------------------------------------
HC2D      IHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSNK
HC2E      IHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSNK
HC2F      ------------------------------------------------------------

HC2A      LLRYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGDN
HC2A-80   ------------------------------------------------------------
HC2B      LLRYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGDN
HC2C      ------------------------------------------------------------
HC2D      LLKYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAVETVVNMLMPHITQKFRDN
HC2E      LLRYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGDN
HC2F      ------------------------------------------------------------

HC2A      PEASKNANHSLAVFIKRCFTFMDRGFVFKQINNYISCFAPGDPKTLFEYKFEFLRVVCNH
HC2A-80   ------------------------------------------------------------
HC2B      PEASKNANHSLAVFIKRCFTFMDRGFVFKQINNYISCFAPGDPKTLFEYKFEFLRVVCNH
HC2C      ------------------------------------------------------------
HC2D      PEASKNANHSLAVFIKRCFTFMDRGFVFKQINNYISCFAPGDPKTLFEYKFEFLRVVCNH
HC2E      PEASKNANHSLAVFIKRCFTFMDRGFVFKQINNYISCFAPGDPKTLFEYKFEFLRVVCNH
HC2F      ------------------------------------------------------------

HC2A      EHYIPLNLPMPFGKGRIQRYQDLQLDYSLTDEFCRNHFLVGLLLREVGTALQEFREVRLI
HC2A-80   ----------------------QLDYSLTDEFCRNHFLVGLLLREVGTALQEFREVRLI
HC2B      EHYIPLNLPMPFGKGRIQRYQDLQLDYSLTDEFCRNHFLVGLLLREVGTALQEFREVRLI
HC2C      ------------------------------------------------------------
HC2D      EHYIPLNLPMPFGKGRIQRYQDLQLDYSLTDEFCRNHFLVGLLLREVGTALQEFREVRLI
HC2E      EHYIPLNLPMPFGKGRIQRYQDLQLDYSLTDEFCRNHFLVGLLLREVGTALQEFREVRLI
HC2F      ------------------------------------------------------------

HC2A      AISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRINVRDVSPFPVNAGMT
HC2A-80   AISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRINVRDVSPFPVNAGMT
HC2B      AISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRINVRDVSPFPVNAGMT
HC2C      ------------------------------------------------------------
HC2D      AISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRINVRDVSPFPVNAGMT
HC2E      AISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRINVRDVSPFPVNAGMT
HC2F      ------------------------------------------------------------

HC2A      VKDESLALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTTSTPNINSVRNADSRGS
HC2A-80   VKDESLALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTTSTPNINSVRNADSRGS
HC2B      VKDESLALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTTSTPNINSVRNADSRGS
HC2C      ------------------------------------------------------------
HC2D      VKDESLALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTTSTPNINSVRNADSRGS
HC2E      VKDESLALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTTSTPNINSVRNADSRGS
HC2F      ----------------------------------------------------ADSRGS
```

FIG. 3B (cont.)

```
HC2A        LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS
HC2A-80     LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS
HC2B        LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS
HC2C        ------------------------------------------------------------
HC2D        LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS
HC2E        LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS
HC2F        LISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSEIKSLLMCFLYILKSMS

HC2A        DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIARNQEGLGPIVHDRKSQTLPVS
HC2A-80     DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIARNQEGLGPIVHDRKSQTLPVS
HC2B        DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIARNQEGLGPIVHDRKSQTLPVS
HC2C        ------------------------------------------------------------
HC2D        DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIAR--------------------
HC2E        DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIARNQEGLGPIVHDRKSQTLPVS
HC2F        DDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIAS-------VR--KISSVLGIS

HC2A        RNRTGMMHARLQQLGSLDNSLTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFT
HC2A-80     RNRTGMMHARLQQLGSLDNSLTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFT
HC2B        RNRTGMMHARLQQLGSLDNSLTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFT
HC2C        ------------------------------------------------------------
HC2D        ---TGMMHARLQQLGSLDNSLTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFT
HC2E        RNRTGMMHARLQQLGSLDNSLTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFT
HC2F        V----------------D-NG-----YGHSDADVLHQSLLEANIATEVCLTALDTLSLFT

HC2A        LAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRA
HC2A-80     LAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRA
HC2B        LAFK--LLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRA
HC2C        ------------------------------------------------------------
HC2D        LAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRA
HC2E        LAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRA
HC2F        LAFKNQLLADHGHNPLMKKK------------------------------------------

HC2A        DMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLI
HC2A-80     DMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLI
HC2B        DMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLI
HC2C        ------------------------------------------------------------
HC2D        DMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLI
HC2E        DMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLI
HC2F        ------------------------------------------------------------

HC2A        ADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHEND
HC2A-80     ADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHEND
HC2B        ADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHEND
HC2C        ------------------------------------------------------------
HC2D        ADVVGIGGTRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHEND
HC2E        ADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHEND
HC2F        ------------------------------------------------------------
```

FIG. 3B (cont.)

```
HC2A      PEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYLTRK
HC2A-80   PEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYLTRK
HC2B      PEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYLTRK
HC2C      ------------------------------------------------------------
HC2D      PEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYLTRK
HC2E      PEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYLTRK
HC2F      ------------------------------------------------------------

HC2A      ----------------------GVFRQGCTAFRVITPNIDEEASMMEDVGMQDVHFNE
HC2A-80   ----------------------GVFRQGCTAFRVITPNIDEEASMMEDVGMQDVHFNE
HC2B      ----------------------GVFRQGCTAFRVITPNIDEEASMMEDVGMQDVHFNE
HC2C      -----------------------FRQGCTAFRVITPNIDEEASMMEDVGMQDVHFNE
HC2D      EAVQWEPPLLPHSHSACLRRSRGGVFRQGCTAFRVITPNIDEEASMMEDVGMQDVHFNE
HC2E      ----------------------GVFRQGCTAFRVITPNIDEEASMMEDVG--------
HC2F      ------------------------------------------------------------

HC2A       DVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRR----------------
HC2A-80    DVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRR----------------
HC2B       DVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRRDFERLAHLYDTLHRAYSK
HC2C       DVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRRDFERLAHLYDTLHRAYSK
HC2D       DVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRRDFERLAHLYDTLHRAYSK
HC2E       ---------------KAERYELIADIYKLIIPIYEKRRDFERLAHLYDTLHRAYSK
HC2F       ------------------------------------------------------------

HC2A      ----------------------------------DFFEDEDGKEYIYKEPKLTPLSE
HC2A-80   ----------------------------------DFFEDEDGKEYIYKEPKLTPLSE
HC2B      VTEVMHSGRRLLGTYFRVAFFG--------------QGFFEDEDGKEYIYKEPKLTPLSE
HC2C      VTEVMHSGRRLLGTYFRVAFFG--------------QGFFEDEDGKEYIYKEPKLTPLSE
HC2D      VTEVMHSGRRLLGTYFRVAFFGQAAQYQFTDSETDVEGFFEDEDGKEYIYKEPKLTPLSE
HC2E      VTEVMHSGRRLLGTYFRVAFFG--------------QGFFEDEDGKEYIYKEPKLTPLSE
HC2F      ------------------------------------------------------------

HC2A      ISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2A-80   ISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2B      ISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2C      ISQRLLKLYSDKFGSENVKMTQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2D      ISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2E      ISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKELQERKTEF
HC2F      ------------------------------------------------------------

HC2A      ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNP
HC2A-80   ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNP
HC2B      ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNP
HC2C      ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPFMYQHHTDLNP
HC2D      ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNP
HC2E      ERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNP
HC2F      ------------------------------------------------------------
```

FIG. 3B (cont.)

```
HC2A        IEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKR
HC2A-80     IEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKR
HC2B        IEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKR
HC2C        IEVHZ-------------------------------------------------------
HC2D        IEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKR
HC2E        IEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKR
HC2F        ------------------------------------------------------------

HC2A        YPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQ
HC2A-80     YPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQ
HC2B        YPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQ
HC2C        ------------------------------------------------------------
HC2D        YPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQ
HC2E        YPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQ
HC2F        ------------------------------------------------------------

HC2A        ICPLEEKTSVLPNSLHIFNAISGTPTSTMVHGMTSSSSVVZ----
HC2A-80     ICPLEEKTSVLPNSLHIFNAISGTPTSTMVHGMTSSSSVVZ----
HC2B        ICPLEEKTSVLPNSLHIFNAISGTPTSTMVHGMTSSSSVVZ----
HC2C        ----------------------------------------
HC2D        LG--------------------------------------
HC2E        ICPLEEKTSVLPNSLHIFNAISGTPTSTMVHGMTSSSSVVZ----
HC2F        ----------------------------------------
```

FIG. 3B (cont.)

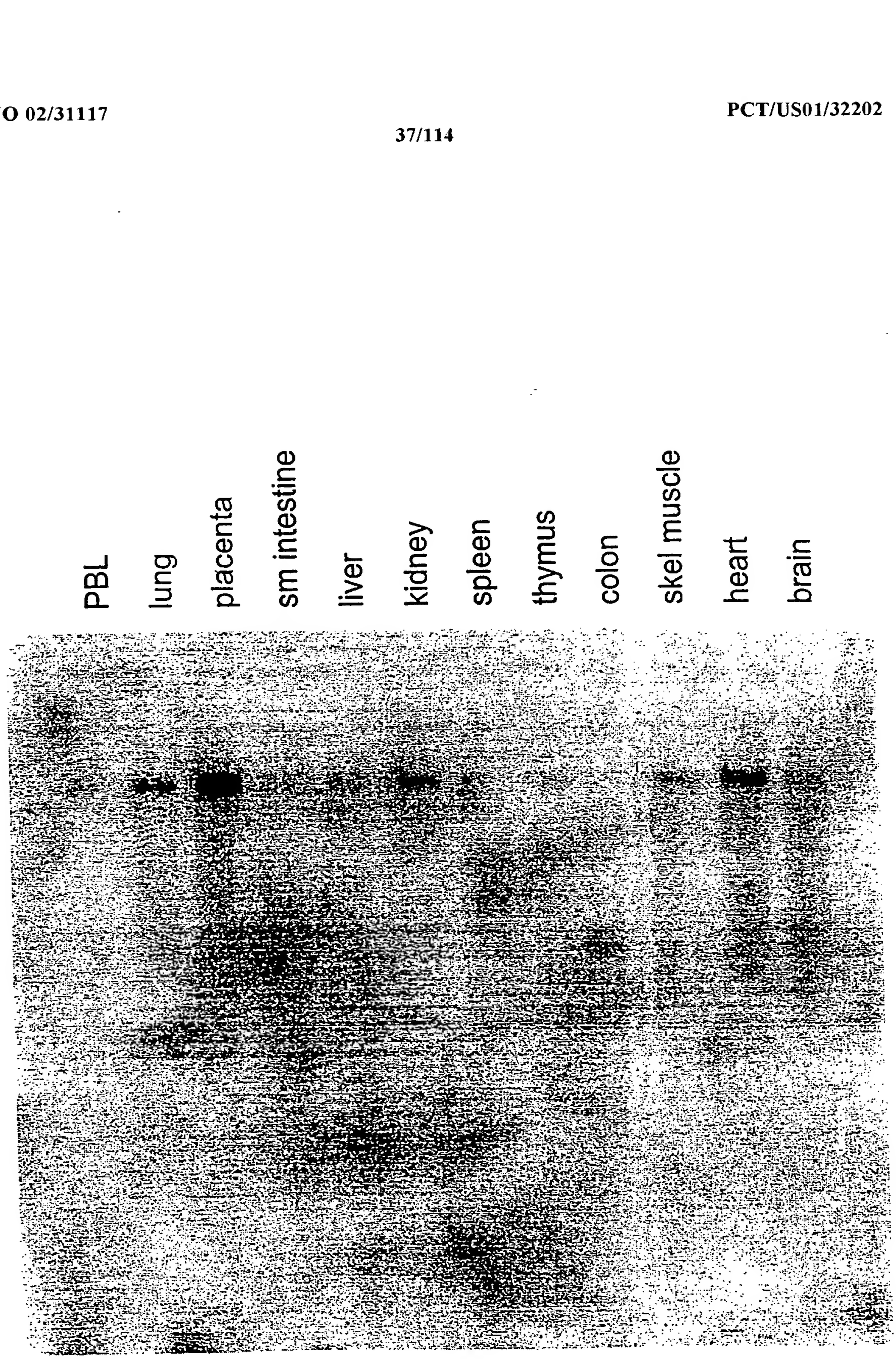


FIG. 4A

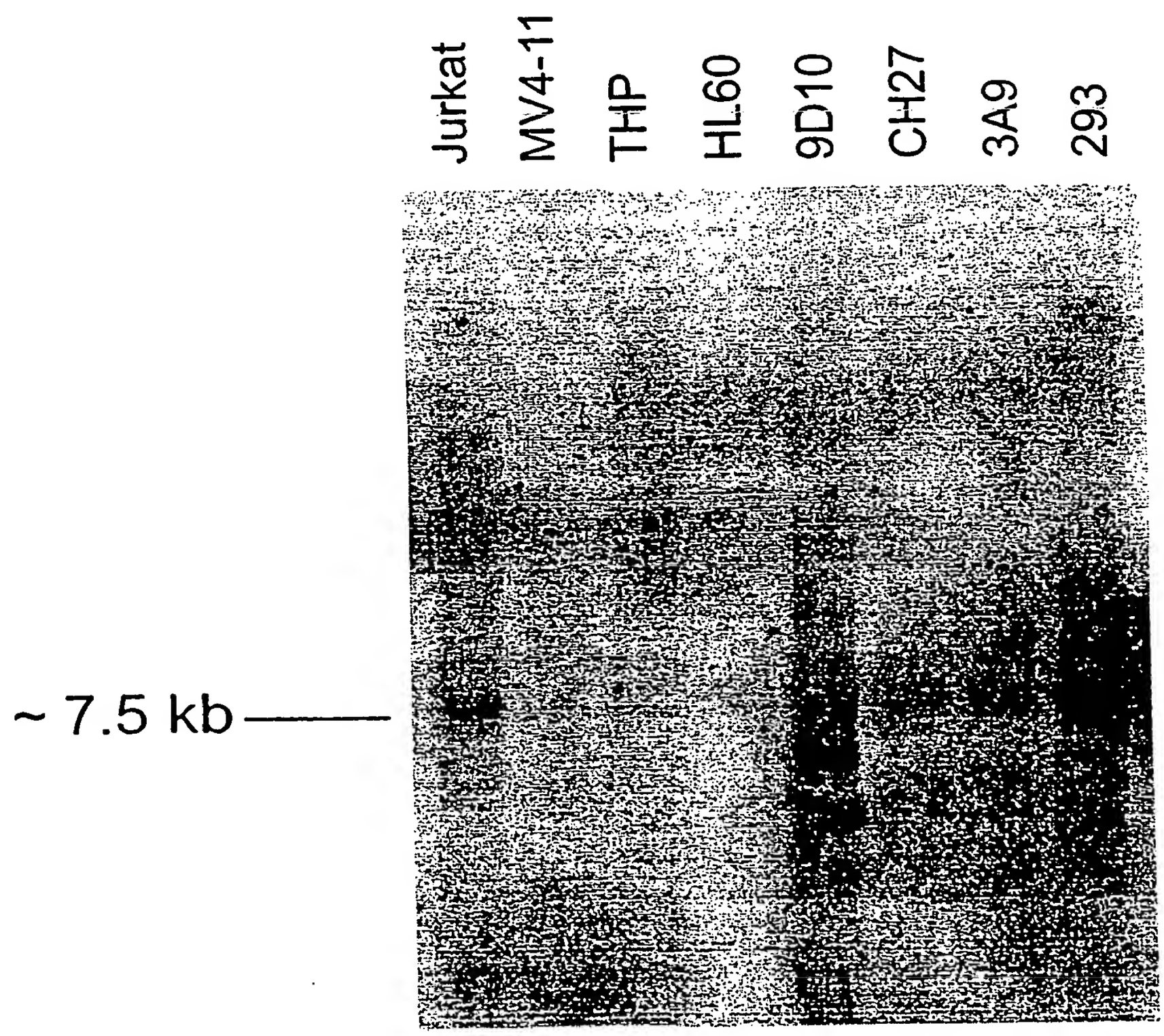


FIG. 4B

```
HC2A    ------------------------------------------------------------
KIAA    ASGNLDKNARFSAIYRQDSNKLSNDDMLKLLADFRKPEKMAKLPVILGNLDITIDNVSSD
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    ------------------------------------------------------------
KIAA    FPNYVNSSYIPTKQFETCSKTPITFEVEEFVPCIPKHTQPYTIYTNHLYVYPKYLKYDSQ
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    ------------------------------------------VLHHHQNPEFYDEIK
KIAA    KSFAKARNIAICIEFKDSDEEDSQPLKCIYGRPGGPVFTRSAFAAVLHHHQNPEFYDEIK
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
KIAA    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
KIAA    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------GPGPARSTVSISLISNSARV
HC5     ------------------------------------------------------------

HC2A    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLT-RATQEEVAVNVTRV
KIAA    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLT-RATQEEVAVNVTRV
rat     ------------------------------------------------------------
HC4     ----------------------ME IQVLIRFLSVILMQLFWVLPNMIHEDDVPISCPMV
HC1     ----------------------------MSFLPIILNQLFKVLV-QNEEDEITTTVTRV
HC3     NRSRSLSNSNPDISGTPTSPDDEVRSIIGSKGLDRSNSWVNTGGPKAAPWGSNPSPSAES
HC5     ------------------------------------------------------------
```

FIG.5A

```
HC2A    IIHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSN
KIAA    IIHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSN
rat     ------------------------------------------------------------
HC4     LFHIVSKCHEEGLDSYLSSFIKYSFRPGKPSAPQAPLIHETLATMMIALLKQSADFLAIN
HC1     LPDIVAKCHEEQLDHSVQSYIKFVFKTR---ACKERPVHEDLAKNVTGLLK-SNDSPTVK
HC3     TQAMDRSCNRMSSHTETSSFLQTLTGRLP----TKKLFHEELALQWVVCSG--SVR---E
HC5     ------------------------------------------------------------

                                        Cadherin
                                        Cleavage
HC2A    KLLRYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGD
KIAA    KLLKYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAVETVVNMLMPHITQKFRD
rat     ------------------------------------------------------------
HC4     KLLKYSWFFFEIIAKSMATYLLEENKIKLTHGQRFPKAYHHALHSLFLAIT-IVESQYAE
HC1     HVLKHSWFFFAIILKSMAQHLIDTNKIQLPRPQRFPESYQNELDNLVMVLSDHVIWKYKD
HC3     SALQQAWFFFELMVKSMVHHLYFNDKLEAPRKSRFPERFMDDIAALVSTIASDIVSRFQK
HC5     ------------------------------------------------------------

HC2A    NPEASKNANHSLAVFIKRCFTFMDRGFVFKQIN---NYIS--CFAPGDPKTLFEYKFEFL
KIAA    NPEASKNANHSLAVFIKRCFTFMDRGFVFKQIN---NYIS--CFAPGDPKTLFEYKFEFL
rat     ------------------------------------------------------------
HC4     IPKESRNVNYSLASFLKCCLTLMDRGFVFNLIN---DYIS--GFSPKDPKVLAEYKFEFL
HC1     ALEETRRATHSVARFLKRCFTFMDRGCVFKMVN---NYIS--MFSSGDLKTLCQYKFDFL
HC3     DTEMVERLNTSLAFFLNDLLSVMDRGFVFSLIKSCYKQVSSKLYSLPNPSVLVSLRLDFL
HC5     ------------------------------------------------------------

HC2A    RVVCNHEHYIPLNLPM-----PFGKGRIQR-----------YQDLQL----DYSLTDEF
KIAA    RVVCNHEHYIPLNLPM-----PFGKGRIQR-----------YQDLQL----DYSLTDEF
rat     ------------------------------------------------------------
HC4     QTICNHEHYIPLNLPM-----AFAKPKLQR-----------VQDSNL----EYSLSDEY
HC1     QEVCQHEHFIPLCLPIRSANIPDPLTPSES-----------TQELHASDMPEYSVTNEF
HC3     RIICSHEHYVTLNLPCSLLTPPASPSPSVSSATSQSSGFSTNVQDQKIANMFELS--VPF
HC5     ---------------MNADTAPTSPCPSIS---SQNSSSCSSFQDQKIASMFDRTSRVPA

HC2A    CRNHFLVGLLLREVGTALQEFRE----VRLIAISVLKNLLIKHSFDDRYASRSHQARIAT
KIAA    CRNHFLVGLLLREVGTALQEFRE----VRLIAISVLKNLLIKHSFDDRYASRSHQARIAT
rat     ------------------------------------------------------------
HC4     CKHHFLVGLLLRETSIALQDNYE----IRYTAISVIKNLLIKHAFDTRYQHKNQQAKIAQ
HC1     CRKHFLIGILLREVGFALQEDQD----VRHLALAVLKNLMAKHSFDDRYREPRKQAQIAS
HC3     RQQHYLAGLVLTELAVILDPDAEGLFGLHKKVINMVHNLLSSHDSDPRYSDPQIKARVAM
HC5     SSTS-SPGLLFTELAAALDAEGEGISEVQRKAVSAIHSLLSSHDLDPRCVKPEVKVKIAA

HC2A    LYLPLFGLLIENVQRINVRDVSPFPVNAG-MTVKDESLALPAVNPLVTPQKGSTLDNSLH
KIAA    LYLPLFGLLIENVQRINVRDVSPFPVNAG-MTVKDESLALPAVNPLVTPQKGSTLDNSLH
rat     ------------------------------------------------------------
HC4     LYLPFVGLLLENIQRLAGRDTLYSCAAMPNSASRDEFPCG-----FTSP--AN--RGSLS
HC1     LYMPLYGMLLDNMPRIYLKDLYPFTVNTSNQGSRDDLSTNGGFQSQTAIKHANSVDTSFS
HC3     LYLPLIGIIMETVPQLYDFTETHNQRGRPICIATDDYESE----------SG---SMIS
HC5     LYLPLVGIILDALPQLCDFTVADTRRYR---TSGSDEEQE----------GA---GAIT

HC2A    KDLLGAISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNSEKSNSLDKHQQSS
KIAA    KDLLGAISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNSEKSNSLDKHQQSS
rat     ------------------------------------------------------------
HC4     TDKDTAYGSFQNG--------HGIKREDSRGSLIP-EGATGFPDQGNTGEN-----TRQS
HC1     KDVLNSIAAFSS------IAISTVNHADSRASLASLDSNPSTNEKSSEKTDNCEKIPRPL
HC3     QTVAMAIAGTSVPQ------------------LTRPGSFLLTSTSGRQHT---------
HC5     QNVALAIAGNNFN--------------------LKTSG-IVLSSLPYKQYN---------
```

FIG. 5A (cont.)

```
HC2A    TLGNSVVRCDKLDQSEIKSLLMCFLYILKSMSDDALFTYWN-KASTSELMDFFTISEVCL
KIAA    TLGNSVVRCDKLDQSEIKSLLMCFLYILKSMSDDALFTYWN-KASTSELMDFFTISEVCL
rat     ------------------------------------------------------------
HC4     STRSSVSQYNRLDQYEIRSLLMCYLYIVKMISEDTLLTYWN-KVSPQELINILILLEVCL
HC1     ALIGSTLRFDRLDQAETRSLLMCFLHIMKTISYETLIAYWQ-RAPSPEVSDFFSILDVCL
HC3     ----------TFSAESSRSLLICLLWVLKN-ADETVLQKWFTDLSVLQLNRLLDLLYLCV
HC5     ----------MLNADTTRNLMICFLWIMKN-ADQSLIRKWIADLPSTQLNRILDLLFICV

HC2A    HQFQYMGKRYIARNQEGLG--PIVHDRKS----------------QTLPVSRNRTGMM
KIAA    HQFQYMGKRYIAR-------------------------------------------TGMM
rat     ------------------------------------------------------------
HC4     FHFRYMGKRNIARVHDAWLSKHFGIDRKS----------------QTMPALRNRSGVM
HC1     QNFRYLGKRNIIRKIAAAF--KFVQSTQNNGTLKGSNPSCQTSGLLAQWMHSTSRHEGHK
HC3     SCFEYKGKKVFERMNSLTFK--KSKDMRAK--------------LEEAILGSIGARQEMV
HC5     LCFEYKGKQSSDKVSTQVLQ--KSRDVKAR--------------LEEALLRGEGARGEMM

HC2A    HARLQQL---------GSLDNS---------LTFNHSYGHSDADVLHQSLLEANIATEVC
KIAA    HARLQQL---------GSLDNS---------LTFNHSYGHSDADVLHQSLLEANIATEVC
rat     ------------------------------------------------------------
HC4     QARLQHL---------SSLESS---------FTLNHSSTTTEADIFHQALLEGNTATEVS
HC1     QHRSQTLPIIRGK---NALSNPKL----LQMLDNTMTSNSNEIDIVHHVDTEANIATEGC
HC3     RRSRGQLERSPSGSAFGSQENLRWRKDMTHWRQNTEKLDKSRAEIEHEALIDGNLATEAN
HC5     RRRAPGNDRFP-----GLNENLRWKKEQTHWRQANEKLDKTKAELDQEALISGNLATEAH

HC2A    LTALDTLSLFTLAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIY
KIAA    LTALDTLSLFTLAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIY
rat     ----------------KLSRGHSPLMKKVFDVYLCFLQKHQSEMALKNVFTALRSLIY
HC4     LTVLDTISFFTQCFKTHFLNNDGHNPLMKKVFDIHLAFLKNGQSEVSLKHVFASLRAFIS
HC1     LTILDLVSLFTQTHQRQLQQCDCQNSLMKRGFDTYMLFFQVNQSATALKHVFASLRLFVC
HC3     LIILDTLEIVVQTVS--VTES--KESILGGVLKVLLHSMACNQSAVYLQHCFATQRALVS
HC5     LIILDMQENIIQASS--ALDC--KDSLLGGVLRVLVNSLNCDQSTTYLTHCFATLRALIA

HC2A    KFPSTFYEGRADMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH
KIAA    KFPSTFYEGRADMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH
rat     KFPSTFYEGRADMCASLCYEVLKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH
HC4     KFPSAFFKGRVNMCAAFCYEVLKCCTSKISSTRNEASALLYLLMRNNFEYTKRKTFLRTH
HC1     KFPSAFFQGPADLCGSFCYEVLKCCNHRSRSTQTEASALLYLFMRKNFEFNKQKSIVRSH
HC3     KFPELLFEEETEQCADLCLRLLRHCSSSIGTIRSHPSASLYLLMRQNFEIGN--NFARVK
HC5     KFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLLMRFSFGATS--NFARVK

HC2A    LQVIISVSQLIADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM
KIAA    LQVIISVSQLIADVVGIGGTRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM
rat     LQVIISLSQLIADVVGIGGTRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM
HC4     LQIIIAVSQLIADVALSGGSRFQESLFIINNFANSDRPMLARAFPAEVKDLTKRIRTVLM
HC1     LQLIKAVSQLIAD-AGIGGSRFQHSLAITNNFANGDKQMKNSNFPAEVKDLTKRIRTVLM
HC3     MQVPMSLSSLVGTSQNFNEEFLRRSLKTILTYAEEDLELRETTFPDQVQDLVFNLHMILS
HC5     MQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTPFPTQVEELLCNLNSILY
```

FIG. 5A (cont.)

```
                                                         Transmembrane
HC2A  ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
KIAA  ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
rat   ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
HC4   ATAQMKEHEKDPEMLIDLQYSLAKSYASTPELRKTWLDSMAKIHVKNGDFSEAAMCYVHV
HC1   ATAQMKEHEKDPEMLVDLQYSLANSYASTPELRRTWLESMAKIHARNGDLSEAAMCYIHI
HC3   DTVKMKEHQEDPEMLIDLMYRIAKGYQTSPDLRLTWLQNMAGKHSERSNHAEAAQCLVHS
HC5   DTVKMREFQEDPEMLMDLMYRIAKSYQASPDLRLTWLQNMAEKHTKKKCYTEAAMCLVHA

                                 SH3
HC2A  TALVAEYLTRKGV----------------------------------FRQGCTAFRVITPN
KIAA  TALVAEYLTRKEA---------VQWEPPLLPHSHSACLRRSRGGVFRQGCTAFRVITPN
rat   TALVAEYLTRKEAD-------LALQREPPVFPYSHTSCQRKSRGGMFRQGCTAFRVITPN
HC4   AALVAEFLHRKKL----------------------------------FPNGCSAFKKITPN
HC1   AALIAEYLKRKGYWKVEKICTASLLSEDTHPCDSNSLLTTPSGGSMFSMGWPAFLSITPN
HC3   AALVAEYLSMLED-------------------------------RKYLPVGCVTFQNISSN
HC5   AALVAEYLSMLED-------------------------------HSYLPVGSVSFQNISSN

HC2A  IDEEASMMEDVGMQD-------VHFNEDVLMELLEQCADGLWKAERYELIADIYKLIIPI
KIAA  IDEEASMMEDVGMQD-------VHFNEDVLMELLEQCADGLWKAERYELIADIYKLIIPI
rat   IDEEASMMEDVGMQD-------VHFNEDVLMELLEQCADGLWKAERLRAGLLTSINSSSP
HC4   IDEEGAMKEDAGMMD-------VHYSEEVLLELLEQCVNGLWKAERYEIISEISKLIGPI
HC1   IKEEGAAKEDSGMHD-------TPYNENILVEQLYMCGEFLWKSERYELIADVNKPIIAV
HC3   VLEESAVSDDVVSPDEEGICSGKYFTESGLVGLLEQAAASFSMAGMYEAVNEVYKVLIPI
HC5   VLEESVVSEDTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPI

                  ITAM   ITAM                ITAM            ITAM
HC2A  YEKRRD------------------------------------------------------
KIAA  YEKRRDFERLAHLYDTLHRAYSKVTEVMHSGRRLLGTYFRVAFFGQAAQYQFTDSETDVE
rat   SMKSGGTLETTHLYDTLHRPYSKVTEVITR-----------A----AGSWDLLPGGLFGQ
HC4   YENRREFENLTQVYRTLHGAYTKILEVMHTKKRLLG--------------TFFRVAFYGQ
HC1   FEKQRDFKKLSDLYYDIHRSYLKVAEVVNSEKRLFG--------------RYYRVAFYGQ
HC3   HEANRDAKKLSTIHGKLQEAFSKIVHQSTGWERMFG--------------TYFRVGFYG-
HC5   LEAHREFRKLTLTHSKLQRAFDSIVNKDH--KRMFG--------------TYFRVGFFG-

                ITAM                ITAM
HC2A  -FFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYA
KIAA  GFFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYA
rat   GFFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKFA
HC4   SFFEEEDGKEYIYKEPKLTGLSEISLRLVKLYGEKFGTENVKIIQDSDKVNAKELDPKYA
HC1   GFFEEEEGKEYIYKEPKLTGLSEISQRLLKLYADKFGADNVKIIQDSNKVNPKDLDPKYA
HC3   TKFGDLDEQEFVYKEPAITKLAEISHRLEGFYGERFGEDVVEVIKDSNPVDKCKLDPNKA
HC5   SKFGDLDEQEFVYKEPAITKLPEISHRLEAFYGQCFGAEFVEVIKDSTPVDKTKLDPNKA

      ITAM
HC2A  YIQVTHVIPFFDEKELQERKTEFERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA
KIAA  YIQVTHVIPFFDEKELQERKTEFERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA
rat   YIQVTHVTPFFDEKELQERKTEFERCHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA
HC4   HIQVTYVKPYFDDKELTERKTEFERNHNISRFVFEAPYTLSGKKQGCIEEQCKRRTILTT
HC1   YIQVTYVTPFFEEKEIEDRKTDFEMHHNINRFVFETPFTLSGKKHGGVAEQCKRRTILTT
HC3   YIQITYVEPYFDTYEMKDRITYFDKNYNLRRFMYCTPFTLDGRAHGELHEQFKRKTILTT
HC5   YIQITFVEPYFDEYEMKDRVTYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTT
```

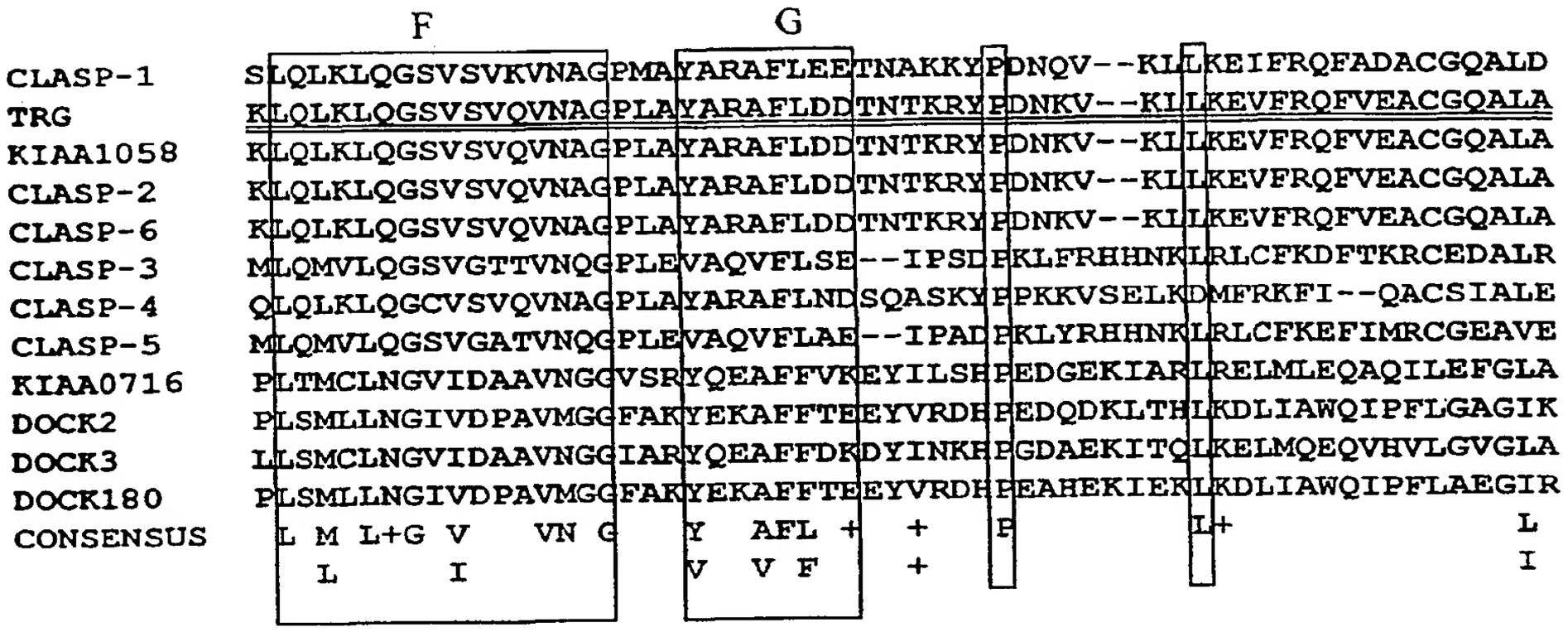
FIG. 5A (cont.)

```
                                                                  Coiled-Coil 1
HC2A    IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSV
KIAA    IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSV
rat     IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELHQLCSSAEVDMIKLQLKLQGSV
HC4     SNSFPYVKKRIPINCEQQINLKPIDGATDEIKDKTAELQKLCSSTDVDMIQLQLKLQGWV
HC1     SHLFPYVKKRIQVISQSSTELNPIEVAIDEMSRKVSELNQLCTMEEVDMISLQLKLQGSV
HC3     SHAFPYIKTRVNVTHKEEIILTPIEVAIEDMQKKTQELAFATHQDPADPKMLQMVLQGSV
HC5     MHAFPYIKTRISVIQKEEFVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSV

                                                                  Coled-Coil 2
HC2A    SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE
KIAA    SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE
rat     SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE
HC4     SVQVNAGPLAYARAFLNDSQASKYPPKKVSELKDMFRKFIQACSIALELNERLIKEDQVE
HC1     SVKVNAGPMAYARAFLEETNAKKYPDNQVKLLKEIFRQFADACGQALDVNERLIKEDQLE
HC3     GTTVNQGPLEVAQVFLSEIPSDPKLFRHHNKLRLCFKDFTKRCEDALRKNKSLIGPVQKE
HC5     GATVNQGPLEVAQVFLAEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQRE

                Coiled-Coil 2
HC2A    YQEEMKANYREMAKELSEIMHEQICPLEEKTS-VLPNSLHIFNAISGTPTSTMVHGMTSS
KIAA    YQEEMKANYREMAKELSEIMHEQLG----------------------------------
rat     YQEEMKANYREIRKELSDIIVPRICPGEDKRATKFPAHLQRHQRDTNKHSGSRVDQFILS
HC4     YHEGLKSNFRDMVKELSDIIHEQILQEDTMHSPWMSNTLHVFCAISGTSSDRGYGSPRYA
HC1     YQEELRSHYKDMLSELSTVMNEQITGRDDLSK---RGVDQTCTRVISKATPALPTVSISS
HC3     YQRELG----KLSS----------PZ----------------------------------
HC5     YQQELKKNYNKLKENLRPMIERKIPELYKPIFRVESQKRDSFHRSSFRKCETQLSQGSZ-

                PBM
HC2A    SSVVZ-------------------------------------------------------
KIAA    -------------------------------------------------------------
rat     CVTLPHEPHVGTCFVMCKLRTTFRANHWFCQAQEEAMGNGREKEPWTVIFNSRFYRSWGK
HC4     EVZ--------------------------------------------------------
HC1     SAEVZ------------------------------------------------------
HC3     -------------------------------------------------------------
HC5     -------------------------------------------------------------

HC2A    -----
KIAA    -----
rat     VHIFF
HC4     -----
HC1     -----
HC3     -----
HC5     -----
```

FIG. 5A (cont.)

```
                    A                                   B

CLASP-1     YRVAFYGQ:::::::::::::GFFEEEEGKEYIYKEP
KIAA1058    FRVAFFGQAAQYQFTDSETDVEGFFEDEDGKEYIYKEP
CLASP-2                             FEDEDGKEYIYKEP
CLASP-6     FRVAFFGQ:::::::::::::GFFEDEDGKEYIYKEP
CLASP-4     FRVAFYGQ:::::::::::::SFFEEEDGKEYIYKEP
DOCK180     FAVGYYGQ::::::GFPTFLRGKVFIYRGKEYERRED
DOCK2       FAVGYYGQ::::::GFPSFLRNKVFIYRGKEYERRED
DOCK3       FRVGFYGR:::::::::::::KFPFFLRNKEYVCRGH
KIAA0716    FRVGFYGK:::::::::::::KFPFFLRNKEFVCRGH
CLASP-3     FRVGFYGT::::::::::::::KFGDLDEQEFVYKEP
CONSENSUS   F V FYG                    KEY   K
                YF                      Q F   R

                                                                   C

TRG         PKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKFAYIQVTHVTPFFDEKE
CLASP-1     PKLTGLSEISQRLLKLYADKFGADNVKIIQDSNKVNPKDLDPKYAYIQVTYVTPFFEEKE
CLASP-2     PKLTPLSEISQRLLKLYSDKFGSENVKMTQDSGKVNPKDLDSKYAYIQVTHVIPFFDEKE
CLASP-4     PKLTGLSEISLRLVKLYGEKFGTENVKIIQDSDKVNAKELDPKYAHIQVTYVKPYFDDKE
CLASP-3     PAITKLAEISHRLEGFYGERFGEDVVEVIKDSNPVDKCKLDPNKAYIQITYVEPYFDTYE
KIAA0716    HDYERLEAFQQRMLNEFPHAIA------MQHANQPDETIFQAEAQYLQIYAVTPIPESQE
DOCK3       HDYERLEAFQQRMLSEFPQAVA------MQHPNHPDDAILQCDAQYLQIYAVTPIPDYVD
DOCK2             FQMQLMTQFPNAEK------MNTTSAPGDDVKNAPGQYIQCFTVQPVLDEHP
DOCK180     EYERREDFQMQLMTQFPNAEK-------MNTTSAPGDDVKNAPGQYIQCFTVQPVLDEHP
CONSENSUS       L         L   Y                            YIQ+   V P  D
                          M   F                             L         E

                  D                          E

CLASP-1     RTILTTSHLFPYVKKRIQVISQSSTELNPIEVAIDEMSRKVSELN
TRG         RTILTAIHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELH
KIAA1058    RTILTAIHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELR
CLASP-2     RTILTAIHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELR
CLASP-6     RTILTAIHCFPYVKKRIPFMYQHHTDLNPIEV:HDEMSKKVAELR
CLASP-4     RTILTTSNSFPYVKKRIPINCEQQINLKPIDVATDEIKDKTAELQ
CLASP-3     KTILTTSHAFPYIKTRVNVTHKEEIILTPIEVAIEDMQKKTQELA
CLASP-5     NTVLTTMHAFPYIKTRISVIQKEEFVLTPIEVAIEDMKKKTLQLA
KIAA0716    RTSLYLVQSLPGISRWFEVEKREVVEMSPLENAIEVLENKNQQLK
DOCK2       RTSFVTAYKLPGILRWFEVVHMSQTTISPLENAIETMSTANEKIL
DOCK3       RTTLTLTHSLPGISRWFEVERRELVEVSPLENAIQVVENKNQELR
DOCK180     RTSFVTAYKLPGILRWFEVVHMSQTTISPLENAIETMSTANEKIL
CONSENSUS   RT L     FP V   + V        + P+E AI+ M      +L
               F     L  L                         +       I
```

FIG. 5B



DOCK2=KIAA0209
DOCK3=KIAA0299
CLASP2variant=KIAA1058

FIG. 5B (cont.)

1
A

```
2                                                   32
GTT TTA CAC CAT CAC CAA AAC CCA GAA TTT TAT GAT GAG ATT AAA ATA GAG TTG CCC ACT
val leu his his his gln asn pro glu phe tyr asp glu ile lys ile glu leu pro thr

62                                                  92
CAG CTG CAT GAA AAG CAC CAC CTG TTG CTC ACA TTC TTC CAT GTC AGC TGT GAC AAC TCA
gln leu his glu lys his his leu leu leu thr phe phe his val ser cys asp asn ser

122                                                 152
AGT AAA GGA AGC ACG AAG AAG AGG GAT GTC GTT GAA ACC CAA GTT GGC TAC TCC TGG CTT
ser lys gly ser thr lys lys arg asp val val glu thr gln val gly tyr ser trp leu

182                                                 212
CCC CTC CTG AAA GAC GGA AGG GTG GTG ACA AGC GAG CAG CAC ATC CCG GTC TCG GCG AAC
pro leu leu lys asp gly arg val val thr ser glu gln his ile pro val ser ala asn

242                                                 272
CTT CCT TCG GGC TAT CTT GGC TAC CAA GAG CTT GGG ATG GGC AGG CAT TAT GGT CCG GAA
leu pro ser gly tyr leu gly tyr gln glu leu gly met gly arg his tyr gly pro glu

302                                                 332
ATT AAA TGG GTA GAT GGA GGC AAG CCA CTG CTG AAA ATT TCC ACT CAT CTG GTT TCT ACA
ile lys trp val asp gly gly lys pro leu leu lys ile ser thr his leu val ser thr
                  ref 1.1, 1.2 and 1.3
362                                                 392
GTG TAT ACT CAG GAT CAG CAT TTA CAT AAT TTT TTC CAG TAC TGT CAG AAA ACC GAA TCT
val tyr thr gln asp gln his leu his asn phe phe gln tyr cys gln lys thr glu ser

422                                                 452
GGA GCC CAA GCC TTA GGA AAC GAA CTT GTA AAG TAC CTT AAG AGT CTG CAT GCG ATG GAA
gly ala gln ala leu gly asn glu leu val lys tyr leu lys ser leu his ala met glu

482                                                 512
GGC CAC GTG ATG ATC GCC TTC TTG CCC ACT ATC CTA AAC CAG CTG TTC CGA GTC CTC ACC
gly his val met ile ala phe leu pro thr ile leu asn gln leu phe arg val leu thr

542                                                 572
AGA GCC ACA CAG GAA GAA GTC GCG GTT AAC GTG ACT CGG GTC ATT ATT CAT GTG GTT GCC
arg ala thr gln glu glu val ala val asn val thr arg val ile ile his val val ala

602                                                 632
CAG TGC CAT GAG GAA GGA TTG GAG AGC CAC TTG AGG TCA TAT GTT AAG TAC GCG TAT AAG
gln cys his glu glu gly leu glu ser his leu arg ser tyr val lys tyr ala tyr lys

662                                                 692
GCT GAG CCA TAT GTT GCC TCT GAA TAC AAG ACA GTG CAT GAA GAA CTG ACC AAA TCC ATG
ala glu pro tyr val ala ser glu tyr lys thr val his glu glu leu thr lys ser met
```

FIG. 6A

```
722                                                 752
ACC ACG ATT CTC AAG CCT TCT GCC GAT TTC CTC ACC AGC AAC AAA CTA CTG AGG TAC TCA
thr thr ile leu lys pro ser ala asp phe leu thr ser asn lys leu leu arg tyr ser

782                                                 812
TGG TTT TTC TTT GAT GTA CTG ATC AAA TCT ATG GCT CAG CAT TTG ATA GAG AAC TCC AAA
trp phe phe phe asp val leu ile lys ser met ala gln his leu ile glu asn ser lys

842              |Cadherin Cleavage|                872
GTT AAG TTG CTG CGA AAC CAG AGA TTT CCT GCA TCC TAT CAT CAT GCA GCG GAA ACC GTT
val lys leu leu arg asn gln arg phe pro ala ser tyr his his ala ala glu thr val

902                                                 932
GTA AAT ATG CTG ATG CCA CAC ATC ACT CAG AAG TTT GGA GAT AAT CCA GAG GCA TCT AAG
val asn met leu met pro his ile thr gln lys phe gly asp asn pro glu ala ser lys

962                                                 992
AAC GCG AAT CAT AGC CTT GCT GTC TTC ATC AAG AGA TGT TTC ACC TTC ATG GAC AGG GGC
asn ala asn his ser leu ala val phe ile lys arg cys phe thr phe met asp arg gly
                                                                    ref 2.1

1022                                                1052
TTT GTC TTC AAG CAG ATC AAC AAC TAC ATT AGC TGT TTT GCT CCT GGA GAC CCA AAG ACC
phe val phe lys gln ile asn asn tyr ile ser cys phe ala pro gly asp pro lys thr

1082                                                1112
CTC TTT GAA TAC AAG TTT GAA TTT CTC CGT GTA GTG TGC AAC CAT GAA CAT TAT ATT CCG
leu phe glu tyr lys phe glu phe leu arg val val cys asn his glu his tyr ile pro

1142                                                1172
TTG AAC TTA CCA ATG CCA TTT GGA AAA GGC AGG ATT CAA AGA TAC CAA GAC CTC CAG CTT
leu asn leu pro met pro phe gly lys gly arg ile gln arg tyr gln asp leu gln leu

1202                                                1232
GAC TAC TCA TTA ACA GAT GAG TTC TGC AGA AAC CAC TTC TTG GTG GGA CTG TTA CTG AGG
asp tyr ser leu thr asp glu phe cys arg asn his phe leu val gly leu leu leu arg

1262                                                1292
GAG GTG GGG ACA GCC CTC CAG GAG TTC CGG GAG GTC CGT CTG ATC GCC ATC AGT GTG CTC
glu val gly thr ala leu gln glu phe arg glu val arg leu ile ala ile ser val leu
                                                                    ref 3.1

1322                                                1352
AAG AAC CTG CTG ATA AAG CAT TCT TTT GAT GAC AGA TAT GCT TCA AGG AGC CAT CAG GCA
lys asn leu leu ile lys his ser phe asp asp arg tyr ala ser arg ser his gln ala

1382                                                1412/471
AGG ATA GCC ACC CTC TAC CTG CCT CTG TTT GGT CTG CTG ATT GAA AAC GTC CAG CGG ATC
arg ile ala thr leu tyr leu pro leu phe gly leu leu ile glu asn val gln arg ile

1442                                                1472
AAT GTG AGG GAT GTG TCA CCC TTC CCT GTG AAC GCG GGC ATG ACC GTG AAG GAT GAA TCC
asn val arg asp val ser pro phe pro val asn ala gly met thr val lys asp glu ser
```

FIG. 6A (cont.)

```
1502                                                1532
CTG GCT CTA CCA GCT GTG AAT CCG CTG GTG ACG CCG CAG AAG GGA AGC ACC CTG GAC AAC
leu ala leu pro ala val asn pro leu val thr pro gln lys gly ser thr leu asp asn
                                                    ref 4.1 and 4.2
1562                                                1592
AGC CTG CAC AAG GAC CTG CTG GGC GCC ATC TCC GGC ATT GCT TCT CCA TAT ACA ACC TCA
ser leu his lys asp leu leu gly ala ile ser gly ile ala ser pro tyr thr thr ser

1622                                                1652
ACT CCA AAC ATC AAC AGT GTG AGA AAT GCT GAT TCG AGA GGA TCT CTC ATA AGC ACA GAT
thr pro asn ile asn ser val arg asn ala asp ser arg gly ser leu ile ser thr asp
                                                    ref 5.1 and 5.2
1682                                                1712
TCG GGT AAC AGC CTT CCA GAA AGG AAT AGT GAG AAG AGC AAT TCC CTG GAT AAG CAC CAA
ser gly asn ser leu pro glu arg asn ser glu lys ser asn ser leu asp lys his gln

1742                                                1772
CAA AGT AGC ACA TTG GGA AAT TCC GTG GTT CGC TGT GAT AAA CTT GAC CAG TCT GAG ATT
gln ser ser thr leu gly asn ser val val arg cys asp lys leu asp gln ser glu ile

1802                                                1832
AAG AGC CTA CTG ATG TGT TTC CTC TAC ATC TTA AAG AGC ATG TCT GAT GAT GCT TTG TTT
lys ser leu leu met cys phe leu tyr ile leu lys ser met ser asp asp ala leu phe

1862                                                1892
ACA TAT TGG AAC AAG GCT TCA ACA TCT GAA CTT ATG GAT TTT TTT ACA ATA TCT GAA GTC
thr tyr trp asn lys ala ser thr ser glu leu met asp phe phe thr ile ser glu val
                    ref 6.1
1922                                                1952
TGC CTG CAC CAG TTC CAG TAC ATG GGG AAG CGA TAC ATA GCC AGG AAC CAG GAG GGG TTG
cys leu his gln phe gln tyr met gly lys arg tyr ile ala arg asn gln glu gly leu

1982                                                2012
GGA CCC ATA GTT CAT GAT CGA AAG TCT CAG ACA TTG CCT GTT TCC CGT AAC AGA ACA GGA
gly pro ile val his asp arg lys ser gln thr leu pro val ser arg asn arg thr gly

2042                                                2072
ATG ATG CAT GCC AGA TTG CAG CAG CTG GGC AGC CTG GAT AAC TCT CTC ACT TTT AAC CAC
met met his ala arg leu gln gln leu gly ser leu asp asn ser leu thr phe asn his

2102                                                2132
AGC TAT GGC CAC TCG GAC GCA GAT GTT CTG CAC CAG TCA TTA CTT GAA GCC AAC ATT GCT
ser tyr gly his ser asp ala asp val leu his gln ser leu leu glu ala asn ile ala

                                                    ref 7.1
2162                                                2192
ACT GAG GTT TGC CTG ACA GCT CTG GAC ACG CTT TCT CTA TTT ACA TTG GCG TTT AAG AAC
thr glu val cys leu thr ala leu asp thr leu ser leu phe thr leu ala phe lys asn

2222                                                2252
CAG CTC CTG GCC GAC CAT GGA CAT AAT CCT CTC ATG AAA AAA GTT TTT GAT GTC TAC CTG
gln leu leu ala asp his gly his asn pro leu met lys lys val phe asp val tyr leu

2282                                                2312
```

FIG. 6A (cont.)

```
TGT TTT CTT CAA AAA CAT CAG TCT GAA ACG GCT TTA AAA AAT GTC TTC ACT GCC TTA AGG
cys phe leu gln lys his gln ser glu thr ala leu lys asn val phe thr ala leu arg

2342                                                2372
TCC TTA ATT TAT AAG TTT CCC TCA ACA TTC TAT GAA GGG AGA GCG GAC ATG TGT GCG GCT
ser leu ile tyr lys phe pro ser thr phe tyr glu gly arg ala asp met cys ala ala

2402                                                2432
CTG TGT TAC GAG ATT CTC AAG TGC TGT AAC TCC AAG CTG AGC TCC ATC AGG ACG GAG GCC
leu cys tyr glu ile leu lys cys cys asn ser lys leu ser ser ile arg thr glu ala

2462                                                2492
TCC CAG CTG CTC TAC TTC CTG ATG AGG AAC AAC TTT GAT TAC ACT GGA AAG AAG TCC TTT
ser gln leu leu tyr phe leu met arg asn asn phe asp tyr thr gly lys lys ser phe

2522                                                2552
GTC CGG ACA CAT TTG CAA GTC ATC ATA TCT GTC AGC CAG CTG ATA GCA GAC GTT GTT GGC
val arg thr his leu gln val ile ile ser val ser gln leu ile ala asp val val gly

2582                                                2612
ATT GGG GAA ACC AGA TTC CAG CAG TCC CTG TCC ATC ATC AAC AAC TGT GCC AAC AGT GAC
ile gly glu thr arg phe gln gln ser leu ser ile ile asn asn cys ala asn ser asp

2642                                                2672
CGG CTT ATT AAG CAC ACC AGC TTC TCC TCT GAT GTG AAG GAC TTA ACC AAA AGG ATA CGC
arg leu ile lys his thr ser phe ser ser asp val lys asp leu thr lys arg ile arg

2702                                                2732
ACG GTG CTA ATG GCC ACC GCC CAG ATG AAG GAG CAT GAG AAC GAC CCA GAG ATG CTG GTG
thr val leu met ala thr ala gln met lys glu his glu asn asp pro glu met leu val

2762                                                2792
GAC CTC CAG TAC AGC CTG GCC AAA TCC TAT GCC AGC ACG CCC GAG CTC AGG AAG ACG TGG
asp leu gln tyr ser leu ala lys ser tyr ala ser thr pro glu leu arg lys thr trp

2822                                                2852          |xxxxxxxxxxxxxxxx Predicted
CTC GAC AGC ATG GCC AGG ATC CAT GTC AAA AAT GGC GAT CTC TCA GAG GCA GCA ATG TGC
leu asp ser met ala arg ile his val lys asn gly asp leu ser glu ala ala met cys

Transmembrane Domain xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx|
TAT GTC CAC GTA ACA GCC CTA GTG GCA GAA TAT CTC ACA CGG AAA GGC GTG TTT AGA CAA
tyr val his val thr ala leu val ala glu tyr leu thr arg lys gly val phe arg gln

2942                                                2972
GGA TGC ACC GCC TTC AGG GTC ATT ACC CCA AAC ATC GAC GAG GAG GCC TCC ATG ATG GAA
gly cys thr ala phe arg val ile thr pro asn ile asp glu glu ala ser met met glu
    ref 8.1
3002                                                3032
GAC GTG GGG ATG CAG GAT GTC CAT TTC AAC GAG GAT GTG CTG ATG GAG CTC CTT GAG CAG
asp val gly met gln asp val his phe asn glu asp val leu met glu leu leu glu gln

3062                                                3092
TGC GCA GAT GGA CTC TGG AAA GCC GAG CGC TAC GAG CTC ATC GCC GAC ATC TAC AAA CTT
cys ala asp gly leu trp lys ala glu arg tyr glu leu ile ala asp ile tyr lys leu
```

FIG. 6A (cont.)

ref 9.1

3122 3152
ATC ATC CCC ATT TAT GAG AAG CGG AGG GAT TTC TTT GAA GAT GAA GAT GGA AAG GAG TAT
ile ile pro ile tyr glu lys arg arg asp phe phe glu asp glu asp gly lys glu tyr

3182 3212
ATT TAC AAG GAA CCC AAA CTC ACA CCG CTG TCG GAA ATT TCT CAG AGA CTC CTT AAA CTG
ile tyr lys glu pro lys leu thr pro leu ser glu ile ser gln arg leu leu lys leu

ref 10.1

3242 3272
TAC TCG GAT AAA TTT GGT TCT GAA AAT GTC AAA ATG ATA CAG GAT TCT GGC AAG GTC AAC
tyr ser asp lys phe gly ser glu asn val lys met ile gln asp ser gly lys val asn

3302 3332
CCT AAG GAT CTG GAT TCT AAG TAT GCA TAC ATC CAG GTG ACT CAC GTC ATC CCC TTC TTT
pro lys asp leu asp ser lys tyr ala tyr ile gln val thr his val ile pro phe phe

3362 3392
GAC GAA AAA GAG TTG CAA GAA AGG AAA ACA GAG TTT GAG AGA TCC CAC AAC ATC CGC CGC
asp glu lys glu leu gln glu arg lys thr glu phe glu arg ser his asn ile arg arg

3422 3452
TTC ATG TTT GAG ATG CCA TTT ACG CAG ACC GGG AAG AGG CAG GGC GGG GTG GAA GAG CAG
phe met phe glu met pro phe thr gln thr gly lys arg gln gly gly val glu glu gln

ref 11.1

3482 3512
TGC AAA CGG CGC ACC ATC CTG ACA GCC ATA CAC TGC TTC CCT TAT GTG AAG AAG CGC ATC
cys lys arg arg thr ile leu thr ala ile his cys phe pro tyr val lys lys arg ile

3542 3572 |xxxxxxxx Coiled-coil 1 xxxxxx
CCT GTC ATG TAC CAG CAC CAC ACT GAC CTG AAC CCC ATC GAG GTG GCC ATT GAC GAG ATG
pro val met tyr gln his his thr asp leu asn pro ile glu val ala ile asp glu met

3602 xxxxxxxx Coiled coil 1 cont'd xxxx 3632 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
AGT AAG AAG GTG GCG GAG CTC CGG CAG CTG TGC TCC TCG GCC GAG GTG GAC ATG ATC AAA
ser lys lys val ala glu leu arg gln leu cys ser ser ala glu val asp met ile lys

ref 12.1

3662 xxxxxxxxxxxxxxxxxxxx | 3692
CTG CAG CTC AAA CTC CAG GGC AGC GTG AGT GTT CAG GTC AAT GCT GGC CCA CTA GCA TAT
leu gln leu lys leu gln gly ser val ser val gln val asn ala gly pro leu ala tyr

3722 3752
GCG CGA GCT TTC TTA GAT GAT ACA AAC ACA AAG CGA TAT CCT GAC AAT AAA GTG AAG CTG
ala arg ala phe leu asp asp thr asn thr lys arg tyr pro asp asn lys val lys leu

3782 3812 |xxxxxxxxxxxxxxxxxx
CTT AAG GAA GTT TTC AGG CAA TTT GTG GAA GCT TGC GGT CAA GCC TTA GCG GTA AAC GAA
leu lys glu val phe arg gln phe val glu ala cys gly gln ala leu ala val asn glu

3842 xxxxxxx Coiled coil 2 xxxxxxxxxx 3872 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
CGT CTG ATT AAA GAA GAC CAG CTC GAG TAT CAG GAA GAA ATG AAA GCC AAC TAC AGG GAA
arg leu ile lys glu asp gln leu glu tyr gln glu glu met lys ala asn tyr arg glu

3902 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx 3932 xxx|

FIG. 6A (cont.)

```
ATG GCG AAG GAG CTT TCT GAA ATC ATG CAT GAG CAG ATC TGC CCC CTG GAG GAG AAG ACG
met ala lys glu leu ser glu ile met his glu gln ile cys pro leu glu glu lys thr

3962                                          3992
AGC GTC TTA CCG AAT TCC CTT CAC ATC TTC AAC GCC ATC AGT GGG ACT CCA ACA AGC ACA
ser val leu pro asn ser leu his ile phe asn ala ile ser gly thr pro thr ser thr

4022                                    |xxxx PBM xxxxx|
ATG GTT CAC GGG ATG ACC AGC TCG TCT TCG GTC GTG TGA TTA CAT CTC ATG GCC CGT GTG
met val his gly met thr ser ser ser ser val val STP

4082                                                  4112
TGG GGA CTT GCT TTG TCA TTT GCA AAC TCA GGA TGC TTT CCA AAG CCA ATC ACT GGG GAG

4142                                                  4172
ACC GAG CAC AGG GAG GAC CAA GGG GAA GGG GAG AGA AAG GAA ATA AAG AAC AAC GTT ATT

4202                                                  4232
TCT TAA CAG ACT TTC TAT AGG AGT TGT AAG AAG GTG CAC ATA TTT TTT TAA ATC TCA CTG

4262                                                  4292
GCA ATA TTC AAA GTT TTC ATT GTG TCT TAA CAA AGG TGT GGT AGA CAC TCT TGA GCT GGA

4322                                                  4352
CTT AGA TTT TAT TCT TCC TTG CAG AGT AGT GTT AGA ATA GAT GGC CTA CAG AAA AAA AAG

4382                                                  4412
GTT CTG GGA TCT ACA TGG CAG GGA GGG CTG CAC TGA CAT TGA TGC CTG GGG GAC CTT TTG

                ref 13.1
4442                                                  4472
CCT CGA CTC GTG CCG GAA ATC TGA TCG TAA TCA GGG TAC AGA ACT TAC TAG TTT TGT CTA

4502                                                  4532
GGA GTA TGT TGT ATG ACT AGG ATT TGT GCT ATT ATC TCA TTC AAC AAC ATA GAG CAA GAA

4562                                                  4592
TAG TGA GCT AAC TGA GCT AGA CAC TCA ATT AAT CCG CTA CTG GCT TCA AGT CAG AAC TTT

                                                      ref 14.1
4622                                                  4652
GTC ATT AAT CAT CGA CTC CGG GAC GGT CAT ATA TGT ATT ACA TTT CTA CAT TTT TAA TAC

4682                                                  4712
TCA CAT GGG CTT ATG CAT TAA GTT TAA TTG TGA TAA ATT TGT GCT GGT CCA GTA TAT GCA

4742                                                  4772
ATA CAC TTT AAT GGT TTA TTC TTG TCA TAA AAA TGT GCA ATA TGG AGA TGT ATA CAA GTC

4802
TTT ACT
```

FIG. 6A (cont.)

BAC sequences of Human CLASP 2

Ref 1.1
Sequence of BAC4 using primer HC2AS2, which spans nucleotides 327-346 of the cDNA. Exon sequence is underlined and represents nucleotides 356-375.
TTCTACAGNGTNTACTCAGGTATGTGCTCCTTCAACAAAATTAGCAGTTGCTGCTCT
GTGACAAAGTTTGCACCATTTTGCAAGAAGAAAAAAATCCTAATGTGTTATATTACTA
TATTTTTACTCTATAGATCTTTTTCTAAAGAAAGAAAGTACAACTGAAGTGCTTATAT
GTATTCATATAAATGACTAGTACAAGCATCATTTTGCAACAGATTTCCCCTTTCATTG
GAGGATCTTCTTGATGTTATTTGTACACGATCAATTTTTAGTCTTAATAAGATGAGGC
TGGGTGTGGTGGCTCACACCTGTAATCCTAGCATTTTGGAGGCCAAGGTGGGCAGAT
CACTTTAGCCCAGGGGTTTGAGACCAGCCTGGCCAACATGGCAAAACCTTGTCTCTA
CAAAAATACNAAAATTATCCAGGCATGGTGATGTGTGCCTGTAGTCCCAACTNCCTAG
GAGGCTAGGGGTAGGGGGATTTGCAAGAGGCTGGGAGGGTCAAAGCCCNAANTGAG
CCATTGGTNCATGTCACTTGGACCCCAAGCNNGGGGNGANCAAGAGCAAAGGACTNN
TGTNNTTTANAAAAAAAACCGGGCTACCATACNNACCAACCCNCNNACCTACCCNACC
TTTCCANNTTAAAANAAGGCTTTGNCTTGCANAGGAAAANCAAAATNNCC

Ref 1.2
Sequence of BAC26 using primer HC2AS2, which spans nucleotides 327-346 of the cDNA. Exon sequence is underlined and represents nucleotides 351-375.
TCTGGTTTCTACAGTGTATACTNAGGTATGTGCTCCTTNAACAAAATTAGCAGTTGCT
GCTCTGTGACAAAGTTTGCACCATTTTGCAAGAAGAAAAAAATCCTAATGTGTTATAT
TACTATATTTTTACTCTATAGATCTTTTTCTAAAGAAAGAAAGTACAACTGAAGTGCTT
ATATGTATTCATATAAATGACTAGTACAAGCATCATTTTGCAACAGATTTCCCCTTTC
ATTGGAGGATCTTCTTGATGTTATTTGTACACGATCAATTTTTAGTCTTAATAAGATG
AGGCTGGGTGTGGTGGCTCACACCTGTAATCCTAGCATTTTGGAGGCCAAGGTGGGC
AGATCACTTTAGCCCAGGGGTTTGAGACCAGCCTGGCCAACATGGCAAAACCTTGTC
TCTACAAAAATACAAAAATTATCCAGGCATGGTGATGTGTGCCTGTAGTCCCAGCTAC
CTAGGAGGCTAGGGTAGGGGGATTGCAAGAGGCTNGGAGGTCAAGGCCCGCAGTGA
GCCATGGTCATGTCACTGCACCCCCAGCCAGGGCCGACAGGAGCAAGACTNTTGTNT
CAAAAAAAAACAGNAACCAACANCCAACAACAACAACNACCTTTCNGCAAAANAAGC
TTGCTNCAANGAAACCAAAATGNCTTCTTNTTTTCCCCCN

Ref 1.3
Sequence of BAC26 using primer HC2AS2, which spans nucleotides 327-346 of the cDNA. Exon sequence is not found within this sequence. This sequence most likely represent intron sequence since this sequence matches the intron sequence found in the previous two BAC sequences.
AGNNNNNCCCNCTACNCCACTTTTAACCTTTTGAAAACACAGTGTTTNCTCAANTATG
CGCTCCTTCACATATTAGCAGTTGCTGCTCTGTGACATAGTTGCACCATTNTGCAAGA
AGAAAAAATCCTAAGTGTNATATCACTATATNNNTACTCTATAGATCTTNTCTAAAGA
AAGAAAGTCAACTGATGTGCTTATATGTATNCATATAAATGACTAGTACATGCATCAT
TTTGCAACAGATNTCTCCTCACATTGGAGGATCTTCTNGANGNATTCGACACGATNAN
TATTAGTCTNAATAAGATGANGCTGGTGTGGNGGTACACTGNATCTAGCATNTGGAN
GCATGTGGCAGACACTTANCCNCGGTNGAGACAGCTGTCACTGNCNAACTGTCTCTN
TAAANCAAANNCTCCGCNGGNGATGGGCTGAGCCAGTCCTAGNNGCTAGNTAGNGAT

FIG. 6A (cont.)

GNNGAGNTGTNGCACGNCGAGNGAGCATGNTCTGTACTGACTCATCAGGCGNCNACA
CGNTCTGTTCNAAAACATACCACACACACTCNCACCTNCGCAAAATTGCTCTNNAAAN
ATGCTTNTTTCACACNGNTNCAATCNCTATATNNTCTTCTATTCTNCNACGTNTNATTA
NNATCTTNCNCTGCANAACNATNCGNCCACCTNNANNACCTTANGCTTNGTTTCACGC
TTATAGCTCCCCTACACNTNNCAGCNNTTNCNNGTGAAGGGCCNCCCGAATCTACGA
NCATACTCTCTCCGTATATNGCCTCGGTCANCGCCATCTGCTGTNTNCTCNTCNCTNG
CNNTTNANCNGTNCGCTATCTCTNNNCCGGATCCNCNCCATATNNTNNCTCTACTTAN
AGCGTAANNTNTNCNCNCACTANTCACAACTTNTNCNTNNAACTCTATCTNCTCCTCT
CTACCACCTCACTTACTACCTNTTCACNCANTCTCCTTCNCTNTCCACTGATCTCCACA
TAGCTGCTNTACTCGCCANTTTATCATATNCACACNCTCTACGCTNNNTNT

Ref 2.1
Sequence of BAC4 using primer HC2S1, which spans nucleotides 1107-1126 of the cDNA. Exon sequence is underlined and represents nucleotides 1079-1097.
CTTGTATTNAAAGAGGGTCTGCAGGAAGAAGTGTGTAGTCATAAATACCTCACTGGA
TATTTTATACAGGATTCTAAAAAACCTATTAGCAATAGTATGCTAGAAATAGTCATTA
GCTTCTTGACCTTCTTAGAACTGCACACTCTATTGCACTGTACAGATTTCAGGATGGC
TGCAGGGATTGATTTGAAAACTAAGGACACATTTCAATAAACAATGTCTTCAATTGAT
TTTTAGGGCTCCTCCTACTTCAATGAAGGACTTCAGGTAGCTTATAATTACAGACACA
GGCTCAATACAATAAAAAAATTAGTAAGGCAGAGCTTTAAAAAAAAAAAAAGGAAAAA
GATAATTCTACCAGAGAAAGGCTACATGGTGACTTCTGTTACCAGTAACAACCCCCCG
CACTACCTTTGGGTCTCCAGGAGCAAAACAGCTAATGTAGTTGTTGATCTGCTTGAAG
ACAAAGCCCCTGTCCATGAAGGTGAAACATCTCTGTGGAGGAAAACAAGCAAAAAAG
TTATTTCAGGTCCAAACATTTCGGAAATTTGGATTCAAAGCAGGCATTTATTGCTAAT
AAGTTTATCCACTGACATAAAAAACATGCCTTCAACATTGCCAGAGCACCTACTCTAT
TNTAGTCNCN

Ref 3.1
Sequence of BAC4 using primer C96AS, which spans nucleotides 1443-1452 of the cDNA. Exon sequence is underlined and represents nucleotides 1370-1422.
AATCAGCAGACCAAACAGAGGCAGGTAGAGGGTGGCTATCCTTGCCTGATGGCTCTG
AAAAGAAGACACACATGGTAAGTTTGACCCAGGATTCTGAGAACCGAACTAAGTTGG
TGCTGACCATCTCCTTTATTTGGATCCTTCCTATAAAGACAGATATTTGATTTTAGTCC
CAAAATAGAGCAAAATCTTAGTGCTGTTACCATGAATTTTCTAACTGATTACTTTCTTT
ACACCACTTAAAATAAAGGACATTATCAATGCACATTCCTTCCATTGGGGACCACTCA
CCCTTGAAGCATATCTGTCATCAAAAGAATGCTTTATCAGCAGGTTCTTGAGCACACT
GATGGCGATCAGACGGACCTCCCGGAACTCCTGGAGGGCTGTCCCCACCTCCCTNAG
TAACAGTCCCACCAAGAAGTGGTTTCTGCAGAACTCATCTGTTAATGAGTAGTCAAGC
TGGGAGGTCTGAAATGAGGATAGAAACTACTTTGNGTTAGGAAAGATGCAATGCTCT
TTTGAATAAAACAAACAAACCAAACNAACAAAAAAAAAAACTAAGACCCATCCTTNTGN
ATTTCAAGCCCACCCTGGGGTNGGTCAAAGAGATGATCAGNANTTTGGCNTTNAAAT
GAAGAAAGAAATNAATTNTCCAGGGGNTGTTCTNCTTTTTAGCACANGGAGGGATNT
TAANTGAAAACCAATTTAAATCCAATTNAGGNG

FIG. 6A (cont.)

Ref 4.1
Sequence of BAC4 using primer C2AS5, which spans nucleotides 1716-1735 of the cDNA. Exon sequence is underlined and represents nucleotides 1602-1703.

TTCCTTTCTGCAAGGCTGTTCCCGAATCTGTGCTTATGAGAGATCCTCTCGAATCAGC
ATTTCTCACACTGTTGATGTTTGGAGTTGAGGTTGTATATGGAGAAGCTAAATGGAAA
TCAAGCCAACAATAAAGTTTTATTAAGACAGAACAAAATAAAGATGAGTACTGAACTT
TAAGGGAAATTGCTTTTATTGCACTTATTTTTTCTGTTAGGAAGTTGGCTCAAGAGTT
GCATTCCATTACTTCACCTTTAAAGAACCAGGTCATATACAATGAGATAAAAAGAAAC
TAGTCTGAAACATTCAGATGTAAACATCAATTCACTTGTTAGAAACCACCTTTGATCG
CTAAAGACTAAATGCATACCTGTTTCAGAATGTGATAGAATGAAGACTTAAAAAAATT
AAAAGATAAATCCACCTACAACTATCAAATCACAAAATTAAACCACACAACAAACTTG
TAGCATTCAAACTGGTAATAAACACTGAGGAGCCTACCCAACTCTGAGGGGTGTCAT
GGGGTATTTTAAATTTTCGAGGAGAACACAGTGATATGTGACCTCAGCCAGAAGCTG
CTGTTTNAGCAGCAGGTTGGTGCTATGCTCCTTTTTGAAGACATATTTGTGAAGCTGG
GTATTTTGGGGGGCCTGCTTATGATAAAANGGCAAGGTNTTCAATGNAGGGGN

Ref 4.2
Sequence of BAC26 using primer C2AS5, which spans nucleotides 1716-1735 of the cDNA. Exon sequence is underlined and represents nucleotides 1602-1703.

TTCCTTTCTGGAAGGCTGTTACCCGAATCTGTGCTTATGAGAGATCCTCTCGAATCAG
CATTTCTCACACTGTTGATGTTTGGAGTTGAGGTTGTATATGGAGAAGCTAAATGGAA
ATCAAGCCAACAATAAAGTTTTATTAAGACAGAACAAAATAAAGATGAGTACTGAACT
TTAAGGGAAATTGCTTTTATTGCACTTATTTTTTCTGTTAGGAAGTTGGCTCAAGAGT
TGCATTCCATTACTTCACCTTTAAAGAACCAGGTCATATACAATGAGATAAAAAGAAA
CTAGTCTGAAACATTCAGATGTAAACATCAATTCACTTGTTAGAAACCACCTTTGATC
GCTAAAGACTAAATGCATACCTGTTTCAGAATGTGATAGAATGAAGACTTAAAAAAAT
TAAAAGATAAATCCACCTACAACTATCAAATCACAAAATTAAACCNCACAACAAACTT
GTAGCATTCAAACTGGTAATAAAACACTGAGGAGCCTACCCAACTTTGAGGGGTGTC
AATGGGGTNTTTTTAAATTTTTCGNGGGANANCCCAGTGNTATGGTGACCTTCACCCA
AGAAGCTTGTTTGTTTNACCAAGCNAGGTTGNNCTNTGCTCCTTTTTAGAANACNNTA
TTTTNNNAAATNCTGGNTTTTTTNNGNGGCCCCCTNCNTTNNT

Ref 5.1
Sequence of BAC4 using primer C2S6, which spans nucleotides 1686-1705 of the cDNA. Exon sequence is underlined and represents nucleotides 1724-1736.

TTCCTGGATAAGGTAATTGCTTTTACCCAACACAAATGTTTCTTATAATCAATGGATT
TAGCCCAAAGTAAACGTACTTCATGTTCTAGTGCCTTTTAAGTGTGACCTTTTGTTTT
TTTCTAAACCACCCGGCTGACCTGGAGTAGGTGATGAGAGCTTTAAGGTTGGGGCCC
ATTCCTTGAAGTGCTCTGATTCCTGTTTCCAGTACCTCAGATCCTGGGCAGGGTTTGC
AGTGGAGCGTCTTGAGTGAATGGCTCTGGTGGGTTGAACGGGGAGGGACTCAAAAT
GCTGCCCATCTCAATTTCCTGTAGTCTTTTTATTTATTTATTTATTTTTTGAGACAGAG
TCTCGCTCTGTCGCCCAGGCTGGAGTACAGCGGCACGATCTCAATTNACTGCAACCT
CCGCCTCC:TGGGTTCAAACGACTCCTCTGCCTCAGCCTCCCCAGCAGC:TGGGACCA
CAGGCACAAGCCACCACCGCCCGGCTAATTTTTTGTNTTTTTAGTA:GAGAT:GGGGTT
TCACCATATTTGGCCAGGCTGGGCTCAAACTCCTGACC:TCGTCATCCGCNCCCTCGG
NCTNCCAAAGTGCTTGGGATTNCAGGCNGTGAGCCCACTTACACCTNGGGCAATTCC
CTGTNAGTCTTTTTTACCAGAGACACCATCATTCAACACAGCTTTTCCACCCACAA

FIG. 6A (cont.)

Ref 5.2
Sequence of BAC26 using primer C2S6, which spans nucleotides 1686-1705 of the cDNA. Exon sequence is underlined and represents nucleotides 1712-1736.

TGAGAAGAGCAATTTCCTGGATAAGGTAATTGCTTTTACCCAACACAAATGTTTCTTA
TAATCAATGGATTTAGCCCAAAGTAAACGTACTTCATGTTCTAGTGCCTTTTAAGTGT
GACCTTTTGTTTTTTTCTAAACCACCCGGCTGACCTGGAGTAGGTGATGAGAGCTTTA
AGGTTGGGGCCCATTCCTTGAAGTGCTCTGATTCCTGTTTCCAGTACCTCAGATCCTG
GGCAGGGTTTGCAGTGGAGCGTCTTGAGTGAATGGCTCTGGTGGGTTGAACGGGGA
GGGACTCAAAATGCTGCCCATCTCAATTTCCTGTAGTCTTTTTATTTATTTATTTATTT
TTTGAGACAGAGTCTCGCTCTGTCGCCCAGGCTGGAGTACAGCGGCACGATCTCAAT
TCACTGCAACCTCCGNCTCCCTGGGTTCAAACGACTCCTCTGNCTNAGNCTCCC:AGC
AGCCTGGGAACCACAGGCTCANGCCACCACGCCCGGCTAATTNTTGTAATTTTNAGT
AANAAATTGGGGGTTCTCACCATNTTGGCCCAAGNCTTGGGCCTAAAAACCTTNCTNA
CCNTCGNCATTCNCNCCCCNACCNTGGGCNCTNCTCAAANGNGCTTGGGGATTTANC
ANNGGCNTTAACCCCCCNTATCACCGTGGNCCTTAATTT

Ref 6.1
Sequence of BAC4 using primer C2S7, which spans nucleotides 1918-1937 of the cDNA. Exon sequence is not found within this sequence.Since the primer is directed against exon sequence we presume that sequence derived from C2S7 is intron sequence.

NAGNGNGGGTTTNAGNCGTTTGAAGCCTGNNACGNGGTGNGTGCTNGAACTCTGTGG
GCTTTCAGGTACTGGGGTATCTGGGAGCCTGCTGTTTGCATTGCTAGTGCATCAGAC
CAGGGCTTTTTCCTCCCTGTAGCTGCTACTTATACACATAGCTCTAACTGAGATGATT
CTCCAGACAACTGATGCAGAGCAGCAAAAGCTTCTGCCGTTCTCCCCTTCTAGGAGT
GTCTCCTTTCTTTGGAAAGAGATCATGAGGGGCTAGATTGTAATGAAGTGAGGCTCA
GTGCTTGAGCACATCCGGTAAAAGTTCCAATATATTGGTCATAAAGTTTCTCATTCTT
TATAGCAGTTAATTTCTCTGGCTCATGAGTTTTCTTAGTTTTAATCTGACTTTTAAATT
AATGTCTCCAGCACCAGTCATATCCCCAGGGCAAACTCAAAGGCATGAGAGGCCAGA
CTCGGGTCCTGGTCATAGCAACCCCTGTCTAGGGCCTTGGTCCCTGCCTCCGCTTGT
GTGCTGTGGCGCAGGTCCTATGGGCCCTTAGGAAACAGGACCACCCTGTCGCACCCC
CTACAGAGACCAGCCAAGTTTGACATTAGATCACCGTAGCAATGTNTGCAAATTCCA
GTTTCTTGCTAAAACAGGTTAAGCCTTGCAGCCACTTTATCTGTAACTGGCNGAGGTT
TTGACATAAAA

Ref 7.1
Sequence of BAC4 using primer C2S8, which spans nucleotides 2143-2162 of the cDNA. Exon sequence is underlined and represents nucleotides 2182-2219.

CTCTCGACACGCTGTTTCTATTAACATTGGCGTTTAAGGTTTGTATCAATTTGCTGTT
CGNGGTTCTAGTTTTACCTTTCACATTCATTCTGCTTGGTAAGCTCAGTGAGCACAAA
CTTACTATGTTGCATTTTTACTTCAGCAATTATTTTTGTCCCTGTAAGGAAACCATTAA
TCTTTAAATTCCTTTAATGAAATCATTCCACAGTGAATGGCTTGAATGCCCTGAAATA
AAATTTAACTGGTCAGTGTGTGCTGCGCGCTTGGGTATGGTGGAAACACGGTCTCTG
GAGGCAGTTAACTCTTGGCTCGAACCTTGAGGATGGTGAATATAGGCACCTAATCAG

FIG. 6A (cont.)

GCATTTCTGCCTTGAATATCTTTAAATATATCCAAATGTTATAGCGTTTAATTAGATTT
TTATGTAGAAAGGAGCAATAAACACAAGACACATGTTTTCAGTTTTTTATCTGTTACT
GCATTAAATGATAAAAACGTTTTGGAGATAGAAAATGAAAGGGGTTTTTTTTTTGTCT
TGTTTTAAAGTTTTAGCAAATAATATTCAAGTAGGTGGAGATGGACTCTTCACCACTC
TCCTGTTTTTAGGAACCCAATACTTTTTCATTCTTGCTAAATGATTACTTCCATTTCTA
GCATAGAAAAGGAGAAAATTGGAATGAGTGTTTATAT

Ref 8.1
Sequence of BAC4 using primer C2S9, which spans nucleotides 2992-3011 of the cDNA. Exon sequence is not found within this sequence. Since the primer is directed against exon sequence we presume that sequence derived from C2S9 is intron sequence

CGCTTTNAAATNCCAGCCGCTACTGCGGGGCGNTNAATTCGAAACGTGTTGTTNTCT
GTGATGCCTGGCTCTGATTGTGTGGGATTGGTCATCAGTGGCGGTTGGCAGNTGGGG
TTCATGGAAGCGGCCATGGGGACTGATGGCAGGCCCTTGGATTGCCACCGCAGAGCC
TGGCAGTGTCTTTGGTCTGCATTCCTACCGGCGAAGTCTCATTTCACCTCACGTGTTA
TCTCTTGGAAAGCATTCCTTTAGCGGGCTGTGTCTACCCTTCCATCCTCTCGTCCAAA
CTCCCCCTCCTTCTCTGTTCTGTCTCCTTCCCATCCTCTTCTCCCCAGTTCTTCTTCCT
ATGTTCCTTCCTCAGTGGTTTCTCTTCCTCTGTTTGACTTTCCAAGGTCATTTTGACTG
TTCCTGCTCCCAACTACAAAGATACTAAAATCTCACCTAACCACTCTTCTTCTTTCTTA
ATGAAAGAATGTTTTCAGTCCATCCCAAATTTGTGTGGACTTCACAAACCTTCTCTAA
AATGGAGCCTTTTCTCTTCCTACTCTTGACTAGNTGGTAAACGCTCCATGTTCTTGGC
CAGAACTCCCTGGTGAGTAGCGTCACTCCCACTTTCCTGTGCAGAACCAAGCCTCCT
AGAAAACTCCTTTGCANCTGAGTGGGTTGGGACACGCCCTTTNTTTGGG

Ref 9.1
Sequence of BAC4 using primer C2AS10, which spans nucleotides 3276-3295 of the cDNA. Exon sequence is underlined and represents nucleotides 3147-3234.

<u>TTTANACCNATNTATCCGNGTCAGTTANAGGAGTCTCTGAGAAATTTCCGACAGCGGT
GTGAGTTTGGGTTCCTTGTAAATATACTCCTTTCCATCTTCATCTTCAAAGAATCCCT</u>
GTGACATAAAGCACAATTAGAGCTATCCCTGAACGTAAGCCCAGGGCTTACCACCTA
GGAAGCGTTCTTTTATTACAAGGGGGAAAAAAAGGAATGGGTCTAAAAATCCAGCTG
AAATGGGCTTTCTGAATGAGAAAGAAAATGCTAATAACATGAAGTCTAGGTGCAAAG
GTAAAGGAAAAACACAACATTGCAAACTTATTCAAGAATGCAGTCATTAAGTGTTGAG
TGAAATGAAAGATTTTGGATACAAGACTAAGCTGTCCCAGGGAAGTCTAATGGGAGT
CAAGCCTGTTTCACTTTCCCAAGAAGCAGAACTCACTANAAAATGATGAGCAGCCCA
CGACAGGCAGGCTCAGAAGTGGACATGCCTCCCTTCTCCTGATGGCTNCCATGCACA
CAGGATTTTATGGCATGAACTGAAGCGTTTGGGGGTCTGGAGTAAGTTTAGTAAAAG
TTAGGTAAAGCTTGTATAAATTGTATTTTTGCTTTACCCGATGAGAAAAAAAATATTN
AAGACCTGGTAGCTTCAATATTCAAGAAAAATATTTTTCATNTCACCCG

FIG. 6A (cont.)

Ref 10.1
Sequence of BAC4 using primer C2S11, which spans nucleotides 3167-3186 of the cDNA. Exon sequence is underlined and represents nucleotides 3231-3296.

NGNANGTGGAGCCNCGANCCAGGGACAATCTNAACCTNCTTAAACTGTACTCGGATN
AATTTGGTTCTGAAAATGTCAAAATGATACAGGATTCTGGCAAGGTATTGACCATGTT
TGGANAAGTTTCATAGCAATGTAATGTTGTGATNCGATTACATATNATATATTTTTAA
ATGTNTATAGAAAAAAACACANGAAAAATATTAAGGATTGTTGGCCCGTGAGTGGCA
GGTGTATNTTCTTNCTGATCCTTTAGNGCTTTCCATTACATGCNTGACATTAAAAAAA
NCTTTATCGCCTAATTTTTGAAACATCTAATTTTACAAAATAATTAACCGTNTGGCCAN
GNATATTNTCATTTTTAGGNCCAGCTATTTAGAAACTCTGACANAAATGAGGGGCTGT
GGCTTNCCTNCCTNNACTTGNCCCTCTTTCNNGNATGTACCACATGAACTTGNCNCCT
CTTTCNNCTNACCGGGTGGCATGTTANAGGACAGGTTGAAACCNCANTNGGGCNGGA
NTTNGGTNNAATTGGGACACAATGGTACNANGCTCTATNGGAATNGAAACTCTCCCN
ACNNNCNGTGNNCCNTGGGGAAAATGNGNCNNATTCATTTTN

Ref 11.1
Sequence of BAC4 using primer C2S12, which spans nucleotides 3474-3493 of the cDNA. Exon sequence is not found within this sequence. Since the primer is directed against exon sequence we presume that sequence derived from C2S9 is intron sequence

AGNANNGTTNNGCAGCTGCANNTCTGGACCCANAGGCCGCANGGGCACGAGCCNGG
ACACGCTCGGCAAAGAGCTGTCCAGAGGGATTCAGAAGCTTCAGGACTGGAAGGGTC
TTTCGAGCTCAGTTAGCCACCCCCACACCCATTTCAGTTTCACATTTATCTAGTGCTT
CCTTTTGAATACTTGGGATGTTTTTCTGTTGATCTGTTGGCACTTCCTTCTTCCACAA
GACCAGAAGCTCATATCCAATCTAAGGTCACTTACCCTTCTGAGAATCTGATGAAAAT
GGCGTGCCTTATGTGCCTAGATGCTTTTGCACACAGTCTAAGGTGACTTATGGACTCC
AGGTCCAGCAGCCACACCCAGTCCTGGGTCTCCGCACAGGGAGGGACCCGTCTTCAC
ACACCTGTCTCAGGTTCTAGCATTGGGCTGCTTCAGCGGTCTCAGGCTGTGAGTAAA
TGGGATGTGAGCTTGGATCGCCCCACGCTGTTGNCCCCCGGGGGGCTTGGCCAGCTG
GCCACTTNGAAATGCCTCCTTTTGCCCAGGAAAGCTCACTGCATTTCAATGGGGNTTN
TCCACGAAGTTCANCTTTANGGG

Ref 12.1
Sequence of BAC4 using primer C2S13, which spans nucleotides 3645-3664 of the cDNA. Exon sequence is underlined and represents nucleotides 3683-3699.

AGNAAGGTNNCTCANTNAANNCAGCGTGAGNGTTCAGGTGAGCCAGGCACAGCAGGC
CGGAGGGCAGCAGGGGACGTCCTTGCCCCTGGGTGACTTGAGAGTCGTTTCCACTAA
CAAGGTCTACTTGAGAGCCTCGGTTTACCAAGTGATCCCTGCTCCCTTCCCCCAACGT
NTGTGACATTTCTCCTGATATCAGAGGGGGAGGAAACCTCATGATCCCTGCCCCCCG
CCCCATGAGGACTGACTGTGGGGACAAAGAGCCAGATCTCATAGACTACCCTGATTT
GTCAGTATTTGGGGAATTCTGGGTGCCTGATTAGAAGCATCAAGACTCTTCTAAATNC
AAAGAAGTGTGGAGAGCAGTAGATTTTCCTATAAAACTGGTGTTGCTGGTTTCTATGA
AAATTGGATCCAAAAAAAGTCCTTAAGTTTACCCTCTTAATGGNATCTTTTGATTAAT
GGAATTCATTATTTTAATATAGCCCAATCAATCCAATTTTTCTTTATTGGTAGCATTTT
TATGTTCTCTTTAAAAAAATCTTGGNCTACCTCCAAAATTTCACAGATGTTCTCCTAG
GGTTTTCCTCCTTTTGGTTCAAGCATCCCATTCAANGTCTTGCAGTCCATTCTGGGG

FIG. 6A (cont.)

Ref 13.1
Sequence of BAC4 using primer C2S14, which spans nucleotides 4289-4308 of the cDNA. Exon sequence is underlined and represents nucleotides 4321-4448.
GACTTANATTTATTCTTCCTTGCAGAGTAGTGTTAGAATAGATGGCCTACAGAAAAAA
AAGGTTCTGGGATCTACATGGCAGGGAGGGCTGCACTGACATTGATGCCTGGGGGAC
CTTTTGCCTCGAGGCTGAGCTGGAAAATCTTGAAAATATTTTTTTTTCCTGTGGCAC
ATTCAGGTTGAATACAAGAACTATTTTTGTGACTATGTTTTTGATGACCTAAGGGAAC
TGACCATTGTAATTTTTGTACCANTGAACCANGAGATTTAAGTGCTTTTATATTCATTT
CCTTGCATTTAAGAAAATATGAAAGCTTAAGGAATTATGTGAGCTTAAAACTAGTCAA
GCANTTTAGAACCAAAGGCCTATNTTNATAACCGCAACTATGCTNAAAAGNACAAAGT
AGTACAGNATATTGNTATGTACATATCATTTGGTAATACACNCCNGGCNTTCTGTACA
TATATGTATTACATTTCTACNTTTTTAATACTCCCNTGGGCTTATGCCNTTAAGGTTAA
NTTGNGATAAATTTNGGCTGTTCCNGTNTATNCNATACNCTTTT

Ref 14.1
Sequence of BAC4 using primer C2AS15, which spans nucleotides 4680-4700 of the cDNA. Exon sequence is underlined and represents nucleotides 4660-4683.
ATGAGAATGTAATACATATATGTACAGAATGCCAGGACTGTATTAACAATGATATGTA
CATAACAATATACTGTACTACTTTGTACTTTTCAGCATAGTTGCGGTTATTAATATAG
GCCTTTGGTTCTAAACTGCTTGACTAGTTTTAAGCTCACATAATTCCTTAAGCTTTCAT
ATTTTCTTAAATGCAAGGAAATGAATATAAAAGCACTAAATCTCCTGGTTCACTGGTA
CAAAAATTACAATGGTCAGTTCCCTTAGGTCATCAAAAACTAGTCACAAAAATAGTTC
TTGTATTCAACCTGAATGTGCCACAGGAAAAAAAAAATATTTTCAAGATTTTCCAGCT
CAGCCTCGAGGCAAAAGGCCCCCAGGCATCAATGTCAGNGCAGCCCTCCTGCCATGT
AGATCCCAGAACCTTTTTTTTCTGTAGGCCATCTATTCTAACACTACTCTGCAGGGAG
AATAAAATCTAAAGNCCAGCTCAAGAGTGCTACCACACCTTTGTTAAGACACAATGAA
AACTTTGGATATTGGCAGGNGAGATTTAAAAAAAAATGTGCCCTTTCTTACCACTCCT
ATAGNAAAGTCTGGTTAAGAAATAACCGTTGGTCTTTATTTTCCTTTTNTTTCCCCTTC
CCTTGGGNCTTCCTGGGGCTCGG

FIG. 6A (cont.)

```
HC2A    ------------------------------------------------------------
KIAA    ASGNLDKNARFSAIYRQDSNKLSNDDMLKLLADFRKPEKMAKLPVILGNLDITIDNVSSD
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    ------------------------------------------------------------
KIAA    FPNYVNSSYIPTKQFETCSKTPITFEVEEFVPCIPKHTQPYTIYTNHLYVYPKYLKYDSQ
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    ----------------------------------------------VLHHHQNPEFYDEIK
KIAA    KSFAKARNIAICIEFKDSDEEDSQPLKCIYGRPGGPVFTRSAFAAVLHHHQNPEFYDEIK
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
KIAA    IELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVETQVGYSWLPLLKDGRVVTSEQHI
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ------------------------------------------------------------
HC5     ------------------------------------------------------------

HC2A    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
KIAA    PVSANLPSGYLGYQELGMGRHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNFFQYC
rat     ------------------------------------------------------------
HC4     ------------------------------------------------------------
HC1     ------------------------------------------------------------
HC3     ----------------------------------------GPGPARSTVSISLISNSARV
HC5     ------------------------------------------------------------

HC2A    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLT-RATQEEVAVNVTRV
KIAA    QKTESGAQALGNELVKYLKSLHAMEGHVMIAFLPTILNQLFRVLT-RATQEEVAVNVTRV
rat     ------------------------------------------------------------
HC4     ----------------------MEIQVLIRFLSVILMQLFWVLPNMIHEDDVPISCPMV
HC1     ----------------------------MSFLPIILNQLFKVLV-QNEEDEITTTVTRV
HC3     NRSRSLSNSNPDISGTPTSPDDEVRSIIGSKGLDRSNSWVNTGGPKAAPWGSNPSPSAES
HC5     ------------------------------------------------------------
```

FIG. 6B

```
                                                                              Ref.
HC2A   IIHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSN
KIAA   IIHVVAQCHEEGLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTTILKPSADFLTSN
rat    ------------------------------------------------------------
HC4    LFHIVSKCHEEGLDSYLSSFIKYSFRPGKPSAPQAPLIHETLATMMIALLKQSADFLAIN
HC1    LPDIVAKCHEEQLDHSVQSYIKFVFKTR---ACKERPVHEDLAKNVTGLLK-SNDSPTVK
HC3    TQAMDRSCNRMSSHTETSSFLQTLTGRLP----TKKLFHEELALQWVVCSG--SVR---E
HC5    ------------------------------------------------------------

                                   Cadherin
                                   Cleavage
HC2A   KLLRYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGD
KIAA   KLLKYSWFFFDVLIKSMAQHLIENSKVKLLRNQRFPASYHHAVETVVNMLMPHITQKFRD
rat    ------------------------------------------------------------
HC4    KLLKYSWFFFEIIAKSMATYLLEENKIKLTHGQRFPKAYHHALHSLFLAIT-IVESQYAE
HC1    HVLKHSWFFFAIILKSMAQHLIDTNKIQLPRPQRFPESYQNELDNLVMVLSDHVIWKYKD   6.1
HC3    SALQQAWFFFELMVKSMVHHLYFNDKLEAPRKSRFPERFMDDIAALVSTIASDIVSRFQK   1.2/1.2/2.1/
HC5    ------------------------------------------------------------

HC2A   NPEASKNANHSLAVFIKRCFTFMDRGFVFKQIN---NYIS--CFAPGDPKTLFEYKFEFL   2.1
KIAA   NPEASKNANHSLAVFIKRCFTFMDRGFVFKQIN---NYIS--CFAPGDPKTLFEYKFEFL
rat    ------------------------------------------------------------
HC4    IPKESRNVNYSLASFLKCCLTLMDRGFVFNLIN---DYIS--GFSPKDPKVLAEYKFEFL
HC1    ALEETRRATHSVARFLKRCFTFMDRGCVFKMVN---NYIS--MFSSGDLKTLCQYKFDFL   7.1
HC3    DTEMVERLNTSLAFFLNDLLSVMDRGFVFSLIKSCYKQVSSKLYSLPNPSVLVSLRLDFL   3.1/3.2
HC5    ------------------------------------------------------------

HC2A   RVVCNHEHYIPLNLPM-----PFGKGRIQR------------YQDLQL----DYSLTDEF
KIAA   RVVCNHEHYIPLNLPM-----PFGKGRIQR------------YQDLQL----DYSLTDEF
rat    ------------------------------------------------------------
HC4    QTICNHEHYIPLNLPM-----AFAKPKLQR------------VQDSNL----EYSLSDEY
HC1    QEVCQHEHFIPLCLPIRSANIPDPLTPSES------------TQELHASDMPEYSVTNEF
HC3    RIICSHEHYVTLNLPCSLLTPPASPSPSVSSATSQSSGFSTNVQDQKIANMFELS--VPF   4.1/4.2
HC5    ---------------MNADTAPTSPCPSIS---SQNSSSCSSFQDQKIASMFDRTSRVPA

HC2A   CRNHFLVGLLLREVGTALQEFRE----VRLIAISVLKNLLIKHSFDDRYASRSHQARIAT   3.1
KIAA   CRNHFLVGLLLREVGTALQEFRE----VRLIAISVLKNLLIKHSFDDRYASRSHQARIAT
rat    ------------------------------------------------------------
HC4    CKHHFLVGLLLRETSIALQDNYE----IRYTAISVIKNLLIKHAFDTRYQHKNQQAKIAQ
HC1    CRKHFLIGILLREVGFALQEDQD----VRHLALAVLKNLMAKHSFDDRYREPRKQAQIAS   8.1
HC3    RQQHYLAGLVLTELAVILDPDAEGLFGLHKKVINMVHNLLSSHDSDPRYSDPQIKARVAM
HC5    SSTS-SPGLLFTELAAALDAEGEGISEVQRKAVSAIHSLLSSHDLDPRCVKPEVKVKIAA

HC2A   LYLPLFGLLIENVQRINVRDVSPFPVNAG-MTVKDESLALPAVNPLVTPQKGSTLDNSLH
KIAA   LYLPLFGLLIENVQRINVRDVSPFPVNAG-MTVKDESLALPAVNPLVTPQKGSTLDNSLH
rat    ------------------------------------------------------------
HC4    LYLPFVGLLLENIQRLAGRDTLYSCAAMPNSASRDEFPCG-----FTSP--AN--RGSLS
HC1    LYMPLYGMLLDNMPRIYLKDLYPFTVNTSNQGSRDDLSTNGGFQSQTAIKHANSVDTSFS   9.1
HC3    LYLPLIGIIMETVPQLYDFTETHNQRGRPICIATDDYESE-----------SG---SMIS
HC5    LYLPLVGIILDALPQLCDFTVADTRRYR---TSGSDEEQE-----------GA---GAIT

                 4.1/4.2
HC2A   KDLLGAISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNSEKSNSLDKHQQSS   5.1/5.2
KIAA   KDLLGAISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNSEKSNSLDKHQQSS
rat    ------------------------------------------------------------
HC4    TDKDTAYGSFQNG--------HGIKREDSRGSLIP-EGATGFPDQGNTGEN-----TRQS
HC1    KDVLNSIAAFSS------IAISTVNHADSRASLASLDSNPSTNEKSSEKTDNCEKIPRPL   10.1
HC3    QTVAMAIAGTSVPQ-------------------LTRPGSFLLTSTSGRQHT---------   3.1
HC5    QNVALAIAGNNFN--------------------LKTSG-IVLSSLPYKQYN---------   2.1
```

FIG. 6B (cont.)

		Ref.
HC2A	TLGNSVVRCDKLDQSEIKSLLMCFLYILKSMSDDALFTYWN-KASTSELMDFFTISEVCL	
KIAA	TLGNSVVRCDKLDQSEIKSLLMCFLYILKSMSDDALFTYWN-KASTSELMDFFTISEVCL	
rat	--	
HC4	STRSSVSQYNRLDQYEIRSLLMCYLYIVKMISEDTLLTYWN-KVSPQELINILILLEVCL	
HC1	ALIGSTLRFDRLDQAETRSLLMCFLHIMKTISYETLIAYWQ-RAPSPEVSDFFSILDVCL	11.1/11.2
HC3	----------TFSAESSRSLLICLLWVLKN-ADETVLQKWFTDLSVLQLNRLLDLLYLCV	
HC5	----------MLNADTTRNLMICFLWIMKN-ADQSLIRKWIADLPSTQLNRILDLLFICV	
HC2A	HQFQYMGKRYIARNQEGLG--PIVHDRKS------------------QTLPVSRNRTGMM	6.1
KIAA	HQFQYMGKRYIAR---TGMM	
rat	--	
HC4	FHFRYMGKRNIARVHDAWLSKHFGIDRKS------------------QTMPALRNRSGVM	
HC1	QNFRYLGKRNIIRKIAAAF--KFVQSTQNNGTLKGSNPSCQTSGLLAQWMHSTSRHEGHK	
HC3	SCFEYKGKKVFERMNSLTFK--KSKDMRAK--------------LEEAILGSIGARQEMV	
HC5	LCFEYKGKQSSDKVSTQVLQ--KSRDVKAR--------------LEEALLRGEGARGEMM	
HC2A	HARLQQL---------GSLDNS---------LTFNHSYGHSDADVLHQSLLEANIATEVC	
KIAA	HARLQQL---------GSLDNS---------LTFNHSYGHSDADVLHQSLLEANIATEVC	
rat	--	
HC4	QARLQHL---------SSLESS---------FTLNHSSTTTEADIFHQALLEGNTATEVS	
HC1	QHRSQTLPIIRGK---NALSNPKL----LQMLDNTMTSNSNEIDIVHHVDTEANIATEGC	12.1/12.2
HC3	RRSRGQLERSPSGSAFGSQENLRWRKDMTHWRQNTEKLDKSRAEIEHEALIDGNLATEAN	6.1/6.2
HC5	RRRAPGNDRFP-----GLNENLRWKKEQTHWRQANEKLDKTKAELDQEALISGNLATEAH	
HC2A	LTALDTLSLFTLAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIY	7.1
KIAA	LTALDTLSLFTLAFKNQLLADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIY	
rat	------------------KLSRGHSPLMKKVFDVYLCFLQKHQSEMALKNVFTALRSLIY	
HC4	LTVLDTISFFTQCFKTHFLNNDGHNPLMKKVFDIHLAFLKNGQSEVSLKHVFASLRAFIS	
HC1	LTILDLVSLFTQTHQRQLQQCDCQNSLMKRGFDTYMLFFQVNQSATALKHVFASLRLFVC	13.1
HC3	LIILDTLEIVVQTVS--VTES--KESILGGVLKVLLHSMACNQSAVYLQHCFATQRALVS	
HC5	LIILDMQENIIQASS--ALDC--KDSLLGGVLRVLVNSLNCDQSTTYLTHCFATLRALIA	3.1
HC2A	KFPSTFYEGRADMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH	
KIAA	KFPSTFYEGRADMCAALCYEILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH	
rat	KFPSTFYEGRADMCASLCYEVLKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTH	
HC4	KFPSAFFKGRVNMCAAFCYEVLKCCTSKISSTRNEASALLYLLMRNNFEYTKRKTFLRTH	
HC1	KFPSAFFQGPADLCGSFCYEVLKCCNHRSRSTQTEASALLYLFMRKNFEFNKQKSIVRSH	
HC3	KFPELLFEEETEQCADLCLRLLRHCSSSIGTIRSHPSASLYLLMRQNFEIGN--NFARVK	7.1/7.2
HC5	KFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLLMRFSFGATS--NFARVK	
HC2A	LQVIISVSQLIADVVGIGETRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM	
KIAA	LQVIISVSQLIADVVGIGGTRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM	
rat	LQVIISLSQLIADVVGIGGTRFQQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLM	
HC4	LQIIIAVSQLIADVALSGGSRFQESLFIINNFANSDRPMLARAFPAEVKDLTKRIRTVLM	
HC1	LQLIKAVSQLIAD-AGIGGSRFQHSLAITNNFANGDKQMKNSNFPAEVKDLTKRIRTVLM	14.1/14.2/15
HC3	MQVPMSLSSLVGTSQNFNEEFLRRSLKTILTYAEEDLELRETTFPDQVQDLVFNLHMILS	
HC5	MQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTPFPTQVEELLCNLNSILY	

FIG. 6B (cont.)

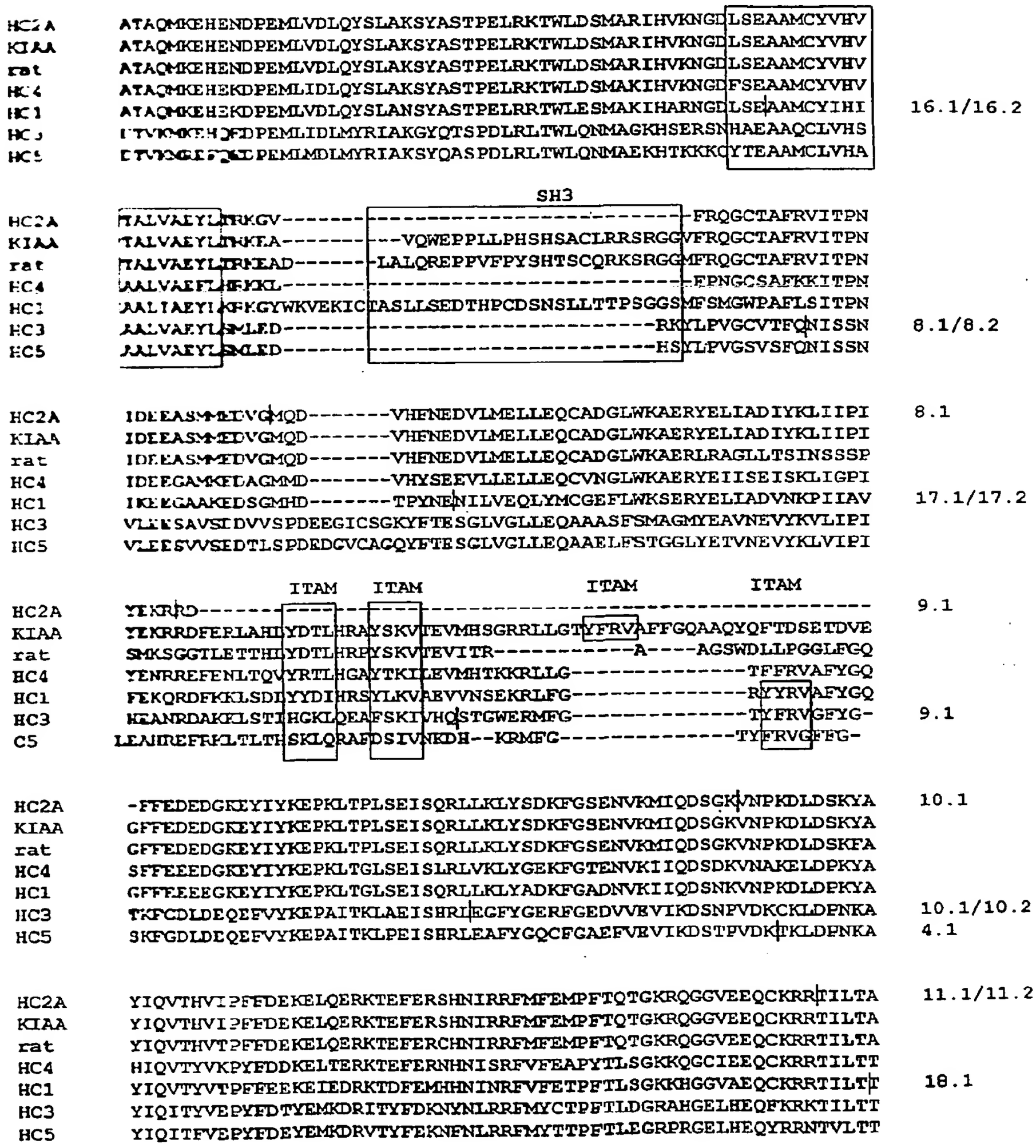

Transmembrane
Ref.
HC2A ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
KIAA ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
rat ATAQMKEHENDPEMLVDLQYSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHV
HC4 ATAQMKEHEKDPEMLIDLQYSLAKSYASTPELRKTWLDSMAKIHVKNGDFSEAAMCYVHV
HC1 ATAQMKEHEKDPEMLVDLQYSLANSYASTPELRRTWLESMAKIHARNGDLSEAAMCYIHI 16.1/16.2
HC3 ITVKMKEHQFDPEMLIDLMYRIAKGYQTSPDLRLTWLQNMAGKHSERSNHAEAAQCLVHS
HC5 DTVKMGEHQFDPEMLMDLMYRIAKSYQASPDLRLTWLQNMAEKHTKKKCYTEAAMCLVHA
SH3
HC2A TALVAEYLHRKGV FRQGCTAFRVITPN
KIAA TALVAEYLHRKEA VQWEPPLLPHSHSACLRRSRGGVFRQGCTAFRVITPN
rat TALVAEYLHRNEAD LALQREPPVFPYSHTSCQRKSRGGMFRQGCTAFRVITPN
HC4 AALVAEFLHEKKL FPNGCSAFKKITPN
HC1 AALIAEYIAFKGYWKVEKICIASLLSEDTHPCDSNSLLTTPSGGSMFSMGWPAFLSITPN
HC3 AALVAEYLAMLED RKYLPVGCVTFQNISSN 8.1/8.2
HC5 AALVAEYLSMLED HSYLPVGSVSFQNISSN
HC2A IDEEASMHEDVGMQD VHFNEDVLMELLEQCADGLWKAERYELIADIYKLIIPI 8.1
KIAA IDEEASMHEDVGMQD VHFNEDVLMELLEQCADGLWKAERYELIADIYKLIIPI
rat IDEEASMHEDVGMQD VHFNEDVLMELLEQCADGLWKAERLRAGLLTSINSSSP
HC4 IDEEGAMKEDAGMMD VHYSEEVLLELLEQCVNGLWKAERYEIISEISKLIGPI
HC1 IKEEGAAKEDSGMHD TPYNENILVEQLYMCGEFLWKSERYELIADVNKPIIAV 17.1/17.2
HC3 VLEESAVSEDVVSPDEEGICSGKYFTESGLVGLLEQAAASFSMAGMYEAVNEVYKVLIPI
HC5 VLEESVVSEDTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPI
ITAM ITAM ITAM ITAM
HC2A YEKRRD 9.1
KIAA YEKRRDFEELAHLYDTLHRAYSKVTEVMHSGRRLLGTYFRVAFFGQAAQYQFTDSETDVE
rat SMKSGGTLETTHLYDTLHRPYSKVTEVITR A AGSWDLLPGGLFGQ
HC4 YENRREFENLTQVYRTLHGAYTKLLEVMHTKKRLLG TFFRVAFYGQ
HC1 FEKQRDFKKLSDLYYDIHRSYLKVAEVVNSEKRLFG RYYRVAFYGQ
HC3 HEANRDAKFLSTLHGKLQEAFSKIVHQSTGWERMFG TYFRVGFYG- 9.1
C5 LEANREFRFLTLTHSKLQRAFDSIVNKDH--KRMFG TYFRVGFFG-
HC2A -FFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYA 10.1
KIAA GFFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYA
rat GFFEDEDGKEYIYKEPKLTPLSEISQRLLKLYSDKFGSENVKMIQDSGKVNPKDLDSKFA
HC4 SFFEEEDGKEYIYKEPKLTGLSEISLRLVKLYGEKFGTENVKIIQDSDKVNAKELDPKYA
HC1 GFFEEEEGKEYIYKEPKLTGLSEISQRLLKLYADKFGADNVKIIQDSNKVNPKDLDPKYA
HC3 TKFCDLDEQEFVYKEPAITKLAEISHRIEGFYGERFGEDVVEVIKDSNFVDKCKLDFNKA 10.1/10.2
HC5 SKFGDLDEQEFVYKEPAITKLPEISHRLEAFYGQCFGAEFVEVIKDSTFVDKTKLDFNKA 4.1
HC2A YIQVTHVIPFFDEKELQERKTEFERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA 11.1/11.2
KIAA YIQVTHVIPFFDEKELQERKTEFERSHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA
rat YIQVTHVTPFFDEKELQERKTEFERCHNIRRFMFEMPFTQTGKRQGGVEEQCKRRTILTA
HC4 HIQVTYVKPYFDDKELTERKTEFERNHNISRFVFEAPYTLSGKKQGCIEEQCKRRTILTT
HC1 YIQVTYVTPFFEEKEIEDRKTDFEMHHNINRFVFETPFTLSGKKHGGVAEQCKRRTILTT 18.1
HC3 YIQITYVEPYFDTYEMKDRITYFDKNYNLRRFMYCTPFTLDGRAHGELHEQFKRKTILTT
HC5 YIQITFVEPYFDEYEMKDRVTYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTT

FIG. 6B (cont.)

```
                                      Coiled-Coil 1                          Ref
HC2A   IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSV
KIAA   IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELRQLCSSAEVDMIKLQLKLQGSV
rat    IHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSKKVAELHQLCSSAEVDMIKLQLKLQGSV
HC4    SNSFPYVKKRIPINCEQQINLKPIDGATDEIKDKTAELQKLCSSTDVDMIQLQLKLQGWV
HC1    SHLFPYVKKRIQVISQSSTELNPIEVAIDEMSRKVSELNQLCTMEEVDMISLQLKLQGSV
HC3    SHAFPYIKTRVNVTHKEEIILTPIEVAIEDMQKKTQELAFATHQDPADPKMLQMVLQGSV 11.1
HC5    MHAFPYIKTRISVIQKEEFVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSV

                                                   Coiled-Coil 2
HC2A   SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE 11.1/12.1
KIAA   SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE
rat    SVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLKEVFRQFVEACGQALAVNERLIKEDQLE
HC4    SVQVNAGPLAYARAFLNDSQASKYPPKKVSELKDMFRKFIQACSIALEINERLIKEDQVE
HC1    SVKVNAGPMAYARAFLEETNAKKYPDNQVKLLKEIFRQFADACGQALDVNERLIKEDQLE
HC3    GTTVNQGPLEVAQVFLSEIPSDPKLFRHHNKLRLCFKDFTKRCEDALRKNKSLIGPVQKE
HC5    GATVNQGPLEVAQVFLAEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQRE

            Coiled-Coil 2
HC2A   YQEEMKANYREMAKELSEIMHEQICPLEEKTS-VLPNSLHIFNAISGTPTSTMVHGMTSS
KIAA   YQEEMKANYREMAKELSEIMHEQLG-----------------------------------
rat    YQEEMKANYREIRKELSDIIVPRICPGEDKRATKFPAHLQRHQRDTNKHSGSRVDQFILS
HC4    YHEGLKSNFRDMVKELSDIIHEQILQEDTMHSPWMSNTLHVFCAISGTSSDRGYGSPRYA
HC1    YQEELRSHYKDMLSELSTVMNEQITGRDDLSK---RGVDQTCTRVISKATPALPTVSISS 19.1
HC3    YQRELG----KLSS----------PZ----------------------------------
HC5    YQQELKKNYNKLKENLRPMIERKIPELYKPIFRVESQKRDSFHRSSFRKCETQLSQGSZ-

       PBM
HC2A   SSVVZ-------------------------------------------------------
KIAA   ------------------------------------------------------------
rat    CVTLPHEPHVGTCFVMCKLRTTFRANHWFCQAQEEAMGNGREKEPWTVIFNSRFYRSWGK
HC4    EVZ---------------------------------------------------------
HC1    SAEVZ-------------------------------------------------------
HC3    ------------------------------------------------------------
HC5    ------------------------------------------------------------

HC2A   -----
KIAA   -----
rat    VHIFF
HC4    -----
HC1    -----
HC3    -----
HC5    -----
```

FIG. 6B (cont.)

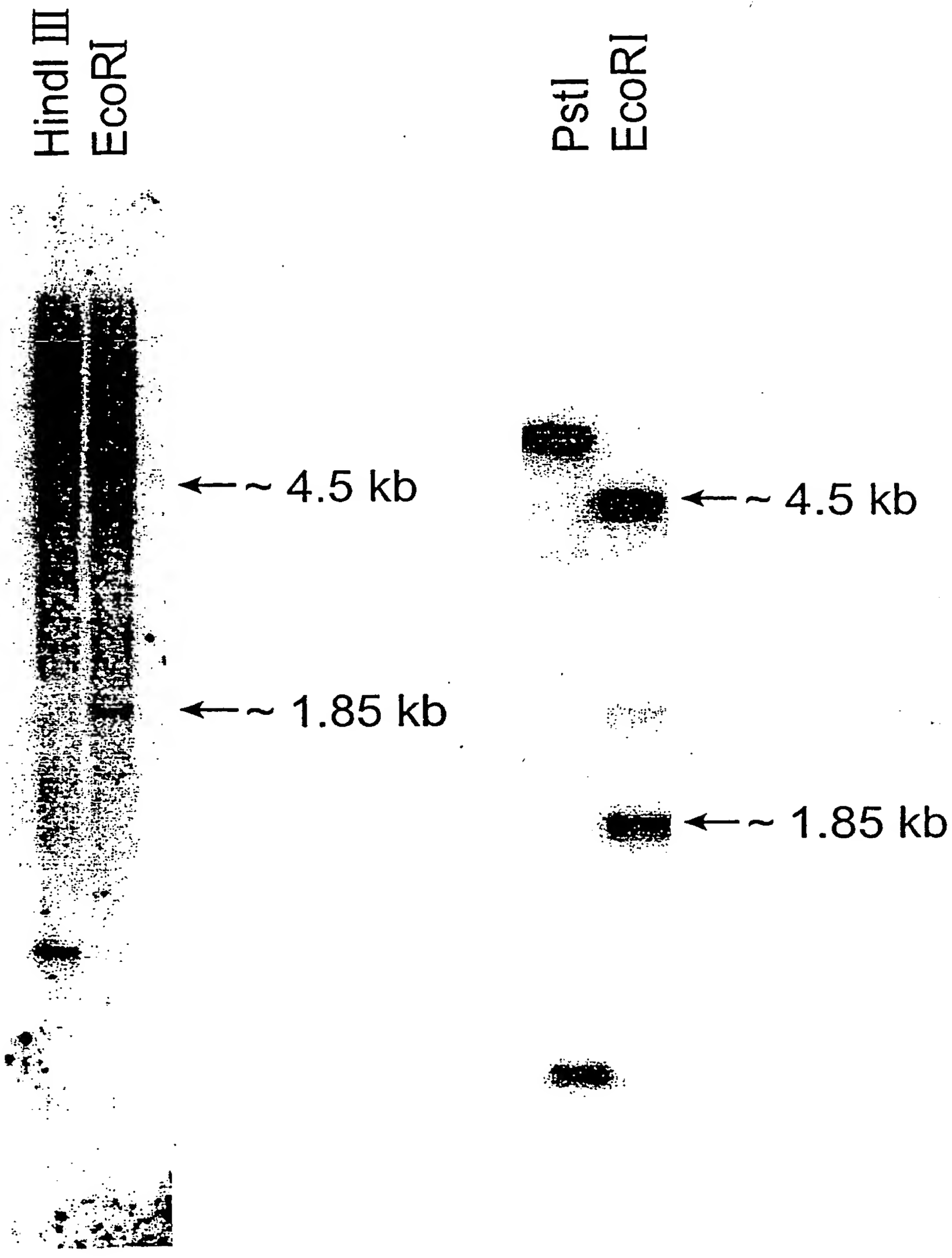

HindI III
EcoRI
~ 4.5 kb
~ 1.85 kb
PstI
EcoRI
~ 4.5 kb
~ 1.85 kb
genomic DNA
BAC 6 DNA

FIG. 7

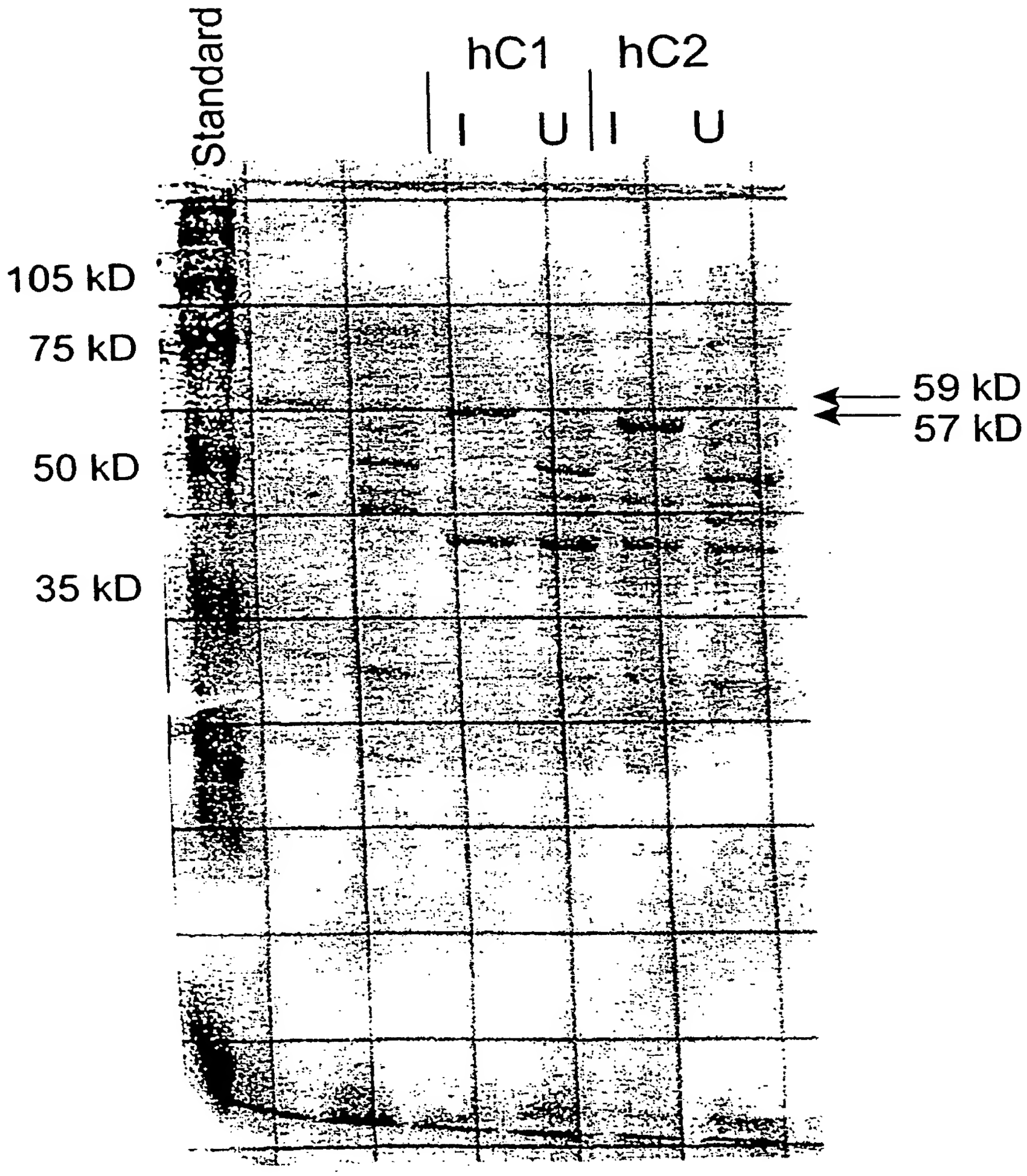


FIG. 8

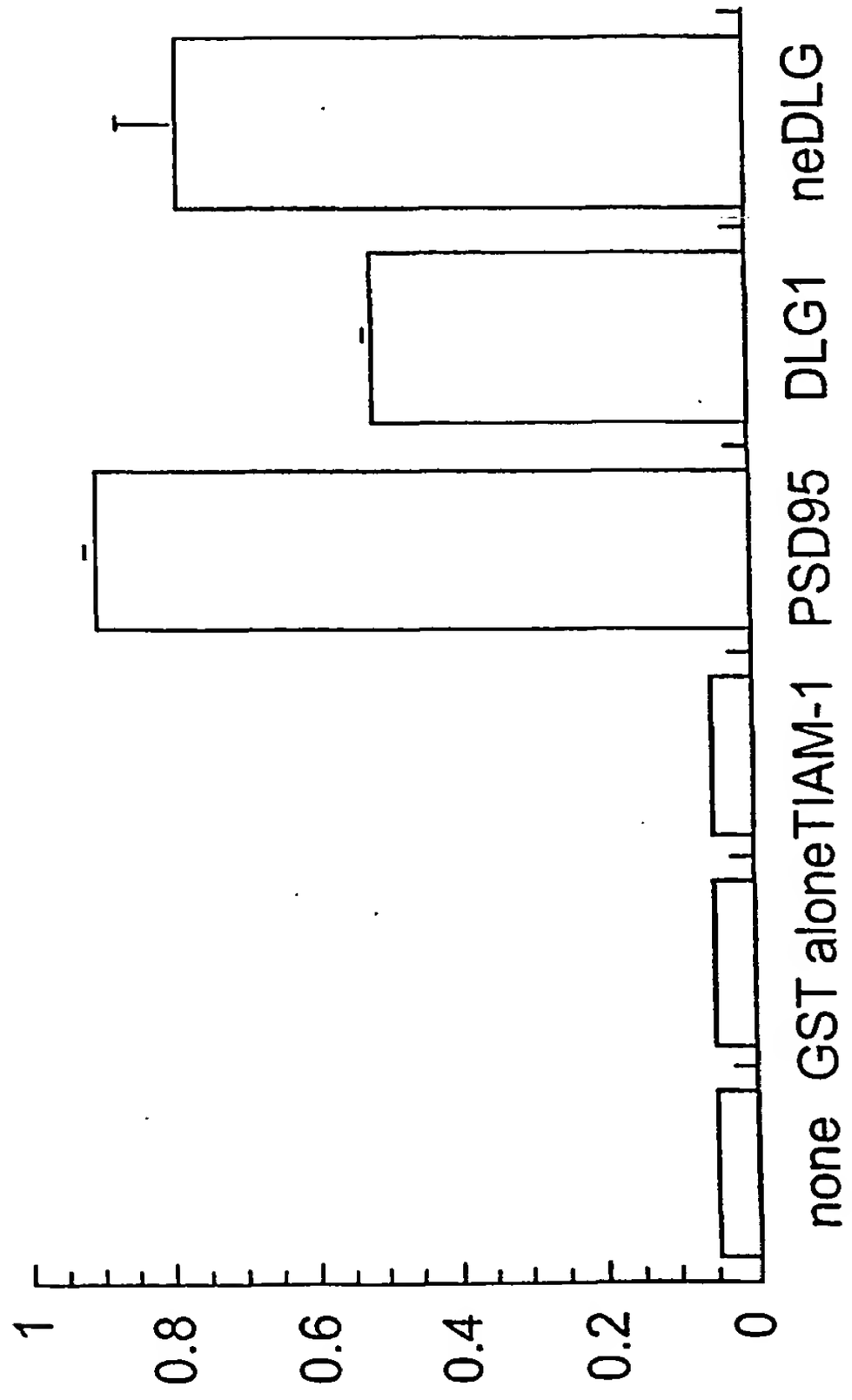


FIG. 9A

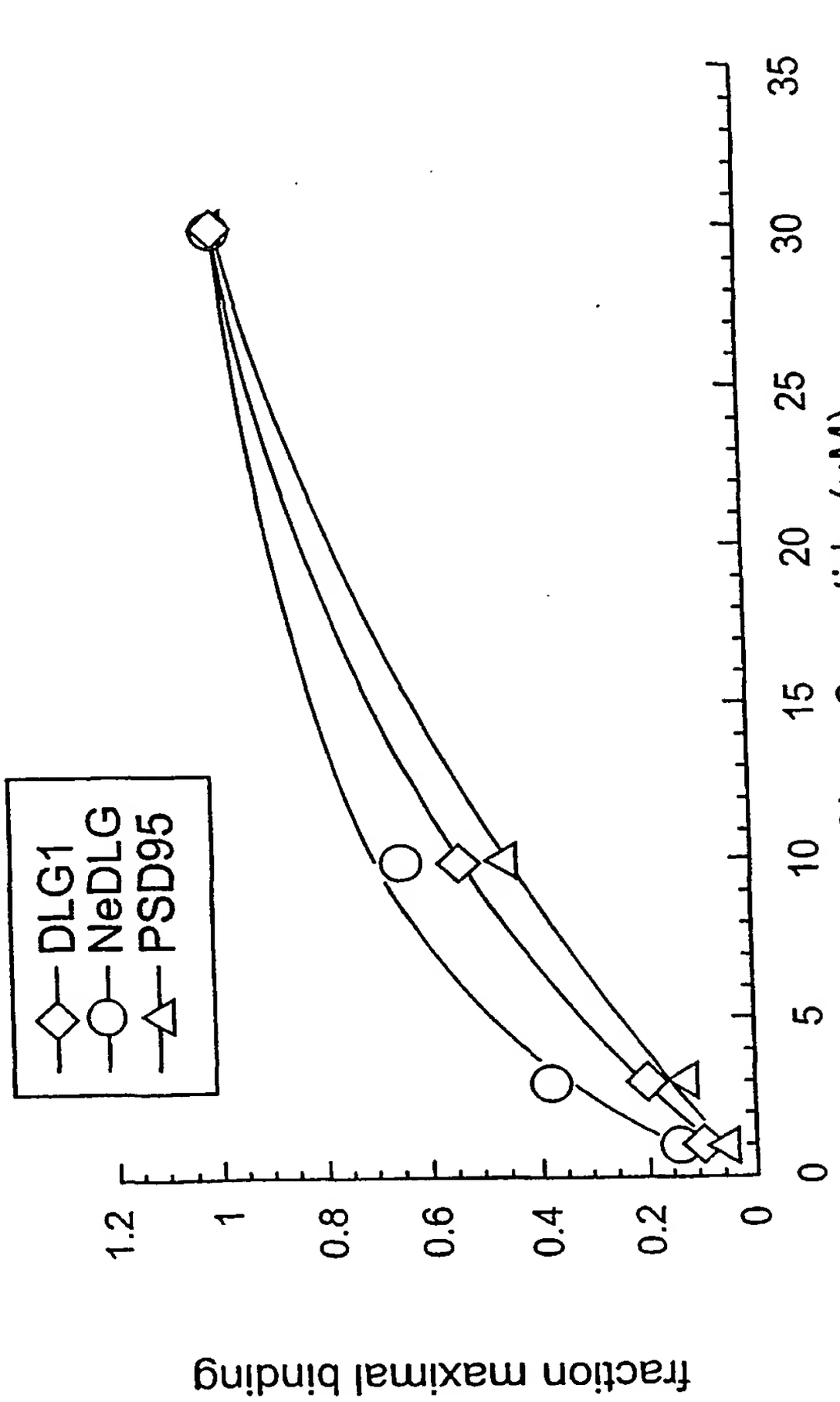

DLG1
NeDLG
PSD95
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
Clasp 2 peptide (uM)
fraction maximal binding

FIG. 9B

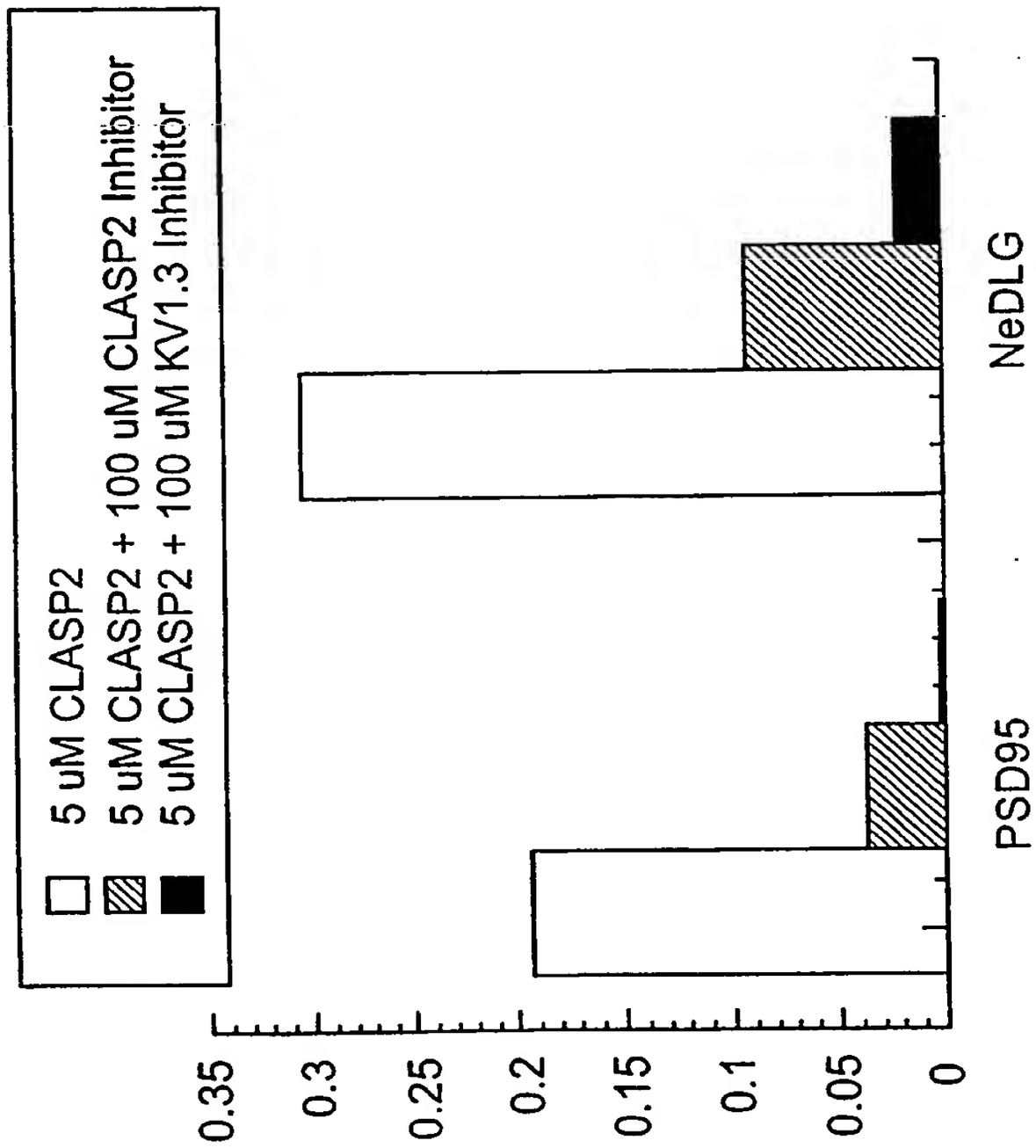


FIG. 9C

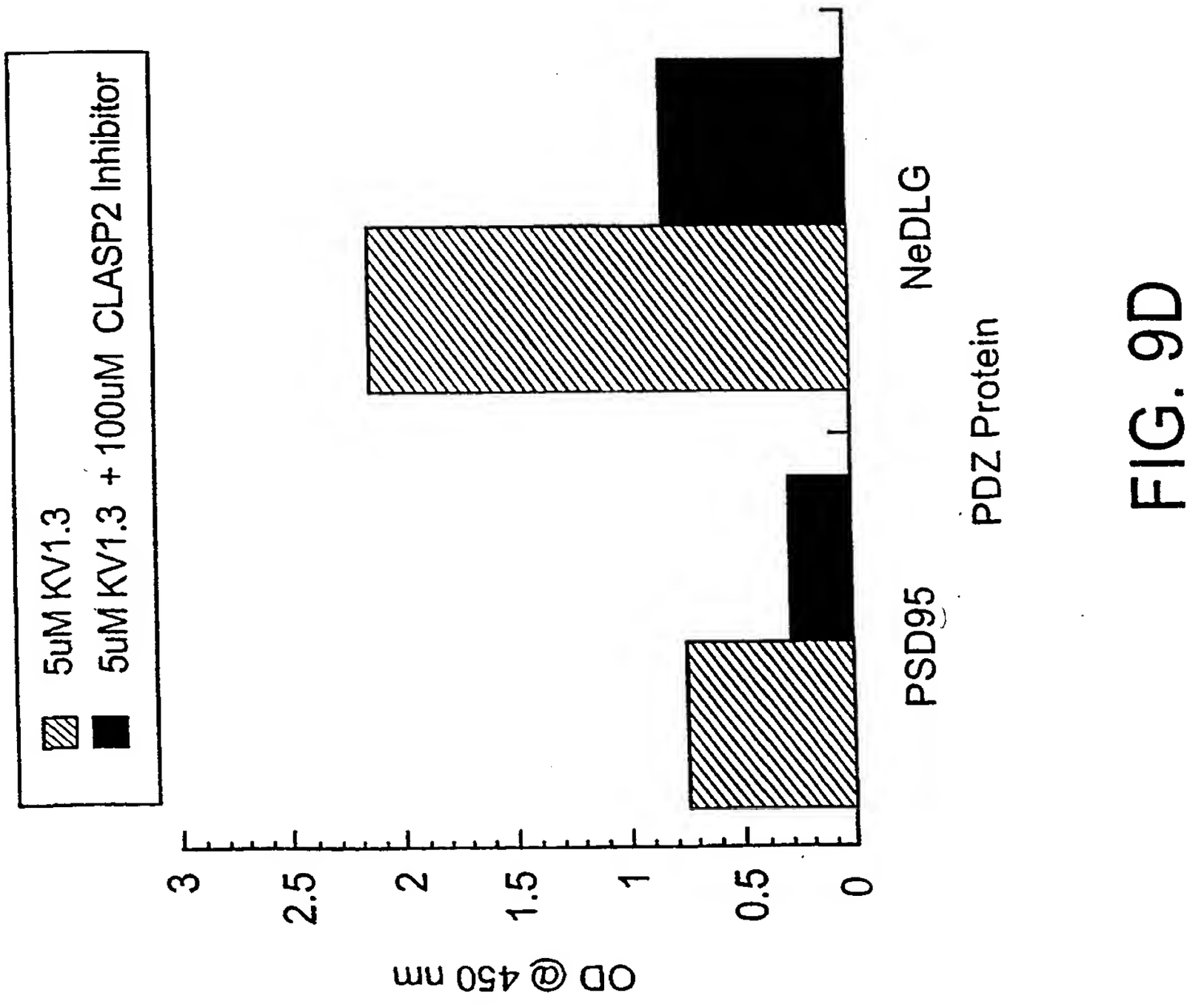

5uM KV1.3
5uM KV1.3 + 100uM CLASP2 Inhibitor
OD @ 450 nm
3
2.5
2
1.5
1
0.5
0
PSD95
NeDLG
PDZ Protein

FIG. 9D

```
     |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
   1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
  81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
 161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
 241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
 321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
 401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
 481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
 561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
 641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
 721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
 801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
 881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
 961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
2321 GCCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT GATGTCTACC TCTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
4161 ATGGCCCCGTG TGTGCGCACT TGCTTTGTCA TTTGCAAACT CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGCAGA 4880
4881 TCTATACAAC TCTTTACT                                                                   4898
     |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10A

```
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
   1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV  VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
  81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN  SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
 161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP  KTLFEYKFEF LKVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
 241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS  VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
 321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL  DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
 401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS  EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
 481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR  TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
 561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV  YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
 641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK  SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
 721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM  LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
 801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM  MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
 881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL  KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
 961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE  EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL  AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE  KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
```

FIG. 10A (cont.)

```
        |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
 1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
 81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA  TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG  CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC  TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA  GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA  TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA  AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA  CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT  CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC  ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT  TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA  GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
 961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT  GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA  TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA  GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT  ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA  GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC  ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC  CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG  TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC  ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA  CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA  GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC  CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG  GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA  AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT  AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG  CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA  CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT  GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT  TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA  ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA  AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA  AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT  CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA  GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG  CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG  AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC  TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA  TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG  ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA  CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA  TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA  CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG  GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT  GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC  TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC  CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA  AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT  ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG  GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA  CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT  CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG  AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT  CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA  GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG  GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG  TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA  TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA  CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT  TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
4881 TGTATACAAG TCTTTACT                                                                    4898
        |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10B

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
    1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV  VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
   81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN  SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
  161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP  KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
  241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS  VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
  321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL  DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
  401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS  EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
  481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR  TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
  561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV  YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
  641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK  SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
  721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM  LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
  801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM  MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
  881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL  KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
  961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE  EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
 1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL  AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
 1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE  KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG. 10B (cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
   1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
  81 GAATTCGCCA CGAGTTTTAC ACCATCACCA AAACCCAGAA  TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
 161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG  CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
 241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC  TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
 321 GGTCTCCGCC AACCTTCCTT CGGGCTATCT TGGCTACCAA  GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
 401 GGGTAGATGC AGGCAAGCCA CTGCTGAAAA TTTCCACTCA  TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
 481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA  AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
 561 CATGCGATGC AAGGCCACGT GATGATCGCC TTCTTGCCCA  CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
 641 ACAGGAAGAA GTCCCGGTTA ACGTGACTCG GGTCATTATT  CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
 721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC  ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
 801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT  TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
 881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA  GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
 961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT  GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA  TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA  GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
1201 GTGTAGTGTC CAACCATGAA CATTATATTC CGTTGAACTT  ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA  GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC  ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
1441 ATGACAGATA TGCTTCAAGC AGCCATCAGG CAAGGATAGC  CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG  TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
1601 ACCAGCTGTC AATCCGCTGC TGACGCCGCA GAAGGGAAGC  ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA  CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA  GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC  CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
1921 TCTTAAAGAC CATGTCTGAT GATGCTTTGT TTACATATTG  GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA  AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT  AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG  CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA  CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT  GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT  TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA  ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA  AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA  AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT  CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA  GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG  CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG  AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC  TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA  TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG  ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA  CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA  TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA  CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG  GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT  GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC  TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC  CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA  AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT  ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG  GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA  CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT  CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG  AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT  CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA  GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG  GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG  TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA  TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA  CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT  TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
4881 TGTATACAAG TCTTTACT                                                                    4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10C

```
          |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
    1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
   81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
  161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
  241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
  321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
  401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS EIKSLIMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
  481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
  561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
  641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
  721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
  801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
  881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
  961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
 1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
 1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
          |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG. 10C (cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
    1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
   81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA  TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
  161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG  CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
  241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC  TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
  321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA  GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
  401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA  TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
  481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA  AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
  561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA  CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
  641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT  CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
  721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC  ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
  801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT  TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
  881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA  GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
  961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT  GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
 1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA  TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
 1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA  GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
 1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT  ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
 1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA  GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
 1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC  ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
 1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC  CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
 1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG  TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
 1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC  ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
 1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA  CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
 1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA  GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
 1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC  CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
 1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG  GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
 2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA  AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
 2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT  AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
 2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG  CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
 2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA  CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
 2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT  GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
 2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT  TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
 2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA  ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
 2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA  AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
 2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA  AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
 2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT  CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
 2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA  GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
 2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG  CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
 2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG  AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
 3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC  TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
 3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA  TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
 3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG  ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
 3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA  CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
 3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA  TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
 3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA  CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
 3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG  GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
 3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT  GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
 3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC  TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
 3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC  CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
 3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA  AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
 3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT  ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
 4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG  GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
 4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA  CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
 4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT  CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
 4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG  AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
 4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT  CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
 4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA  GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
 4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG  GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
 4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG  TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
 4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA  TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
 4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA  CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
 4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT  TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
 4881 TGTATACAAG TCTTTACT                                                                    4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10D

```
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
   1 MEGHVMLAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV  VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
  81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN  SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
 161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP  KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
 241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS  VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
 321 RINVRDVSPF PVNAQNTVKD ESLALPAVNP LVTPQKGSTL  DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
 401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS  EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
 481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR  TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
 561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV  YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
 641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK  SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
 721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM  LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
 801 MCYVHVTALV AEYLTRKGVT RQGCTAFRVI TPNIDEEASM  MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
 881 KLIIPIYEKR KDFFEDEDGK EYIYKEPKLT PLSEISQRLL  KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
 961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE  EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL  AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE  KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
```

FIG. 10D(cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
   1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
  81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA  TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
 161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG  CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
 241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC  TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
 321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA  GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
 401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA  TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
 481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA  AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
 561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA  CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
 641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT  CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
 721 ACTTCAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC  ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
 801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT  TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
 881 CTTTCATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA  GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
 961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT  GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA  TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA  GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT  ACCAATGCCA TTTGGAAAAC GCAGGATTCA AAGATACCAA 1280
1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA  GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC  ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC  CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG  TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC  ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA  CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA  GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC  CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG  GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA  AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT  AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG  CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA  CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT  GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT  TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA  ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA  AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA  AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT  CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA  GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG  CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG  AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC  TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA  TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG  ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA  CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA  TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA  CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG  GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT  GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC  TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC  CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA  AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT  ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG  GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA  CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT  CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG  AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT  CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA  GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG  GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG  TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA  TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA  CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT  TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
4881 TGTATACAAG TCTTTACT                                                                    4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10E

```
        |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
   1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
  81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
 161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
 241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
 321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
 401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
 481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
 561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
 641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
 721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
 801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
 881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
 961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
        |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG. 10E (cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
   1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
  81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
 161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
 241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
 321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
 401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
 481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
 561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
 641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
 721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
 801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
 881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
 961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
4881 TGTATACAAG TCTTTACT                                                                    4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10F

```
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
   1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV  VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
  81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN  SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
 161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP  KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
 241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS  VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
 321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL  DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
 401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS  EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
 481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR  TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
 561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV  YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
 641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK  SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
 721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM  LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
 801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM  MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
 881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL  KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
 961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE  EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL  AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE  KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
```

FIG. 10F (cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
    1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
   81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
  161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
  241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
  321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
  401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
  481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
  561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
  641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
  721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
  801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
  881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
  961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
 1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
 1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
 1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
 1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
 1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
 1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
 1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
 1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
 1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
 1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
 1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
 1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
 2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
 2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
 2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
 2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
 2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
 2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
 2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
 2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
 2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
 2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
 2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
 2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
 2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
 3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
 3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
 3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
 3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
 3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
 3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
 3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
 3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
 3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
 3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
 3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
 3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
 4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
 4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
 4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
 4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
 4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
 4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
 4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
 4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
 4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
 4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
 4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
 4881 TGTATACAAG TCTTTACT                                                                  4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10G

```
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
    1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
   81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
  161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
  241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
  321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
  401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS EIKSLIMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
  481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
  561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
  641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
  721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
  801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
  881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
  961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
 1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
 1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
```

FIG. 10G (cont.)

```
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
    1 AATTGTAATA CGACTCACTA TAGGGCGAAT TGGGTACCGG  GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATATC 80
   81 GAATTCGGCA CGAGTTTTAC ACCATCACCA AAACCCAGAA  TTTTATGATG AGATTAAAAT AGAGTTGCCC ACTCAGCTGC 160
  161 ATGAAAAGCA CCACCTGTTG CTCACATTCT TCCATGTCAG  CTGTGACAAC TCAAGTAAAG GAAGCACGAA GAAGAGGGAT 240
  241 GTCGTTGAAA CCCAAGTTGG CTACTCCTGG CTTCCCCTCC  TGAAAGACGG AAGGGTGGTG ACAAGCGAGC AGCACATCCC 320
  321 GGTCTCGGCG AACCTTCCTT CGGGCTATCT TGGCTACCAA  GAGCTTGGGA TGGGCAGGCA TTATGGTCCG GAAATTAAAT 400
  401 GGGTAGATGG AGGCAAGCCA CTGCTGAAAA TTTCCACTCA  TCTGGTTTCT ACAGGGATAC TCAGGATCAG CATTTACATA 480
  481 ATTTTTTCCA GTACTGTCAG AAAACCGAAT CTGGAGCCCA  AGCCTTAGGA AACGAACTTG TAAAGTACCT TAAGAGTCTG 560
  561 CATGCGATGG AAGGCCACGT GATGATCGCC TTCTTGCCCA  CTATCCTAAA CCAGCTGTTC CGAGTCCTCA CCAGAGCCAC 640
  641 ACAGGAAGAA GTCGCGGTTA ACGTGACTCG GGTCATTATT  CATGTGGTTG CCCAGTGCCA TGAGGAAGGA TTGGAGAGCC 720
  721 ACTTGAGGTC ATATGTTAAG TACGCGTATA AGGCTGAGCC  ATATGTTGCC TCTGAATACA AGACAGTGCA TGAAGAACTG 800
  801 ACCAAATCCA TGACCACGAT TCTCAAGCCT TCTGCCGATT  TCCTCACCAG CAACAAACTA CTGAGGTACT CATGGTTTTT 880
  881 CTTTGATGTA CTGATCAAAT CTATGGCTCA GCATTTGATA  GAGAACTCCA AAGTTAAGTT GCTGCGAAAC CAGAGATTTC 960
  961 CTGCATCCTA TCATCATGCA GCGGAAACCG TTGTAAATAT  GCTGATGCCA CACATCACTC AGAAGTTTGG AGATAATCCA 1040
 1041 GAGGCATCTA AGAACGCGAA TCATAGCCTT GCTGTCTTCA  TCAAGAGATG TTTCACCTTC ATGGACAGGG GCTTTGTCTT 1120
 1121 CAAGCAGATC AACAACTACA TTAGCTGTTT TGCTCCTGGA  GACCCAAAGA CCCTCTTTGA ATACAAGTTT GAATTTCTCC 1200
 1201 GTGTAGTGTG CAACCATGAA CATTATATTC CGTTGAACTT  ACCAATGCCA TTTGGAAAAG GCAGGATTCA AAGATACCAA 1280
 1281 GACCTCCAGC TTGACTACTC ATTAACAGAT GAGTTCTGCA  GAAACCACTT CTTGGTGGGA CTGTTACTGA GGGAGGTGGG 1360
 1361 GACAGCCCTC CAGGAGTTCC GGGAGGTCCG TCTGATCGCC  ATCAGTGTGC TCAAGAACCT GCTGATAAAG CATTCTTTTG 1440
 1441 ATGACAGATA TGCTTCAAGG AGCCATCAGG CAAGGATAGC  CACCCTCTAC CTGCCTCTGT TTGGTCTGCT GATTGAAAAC 1520
 1521 GTCCAGCGGA TCAATGTGAG GGATGTGTCA CCCTTCCCTG  TGAACGCGGG CATGACCGTG AAGGATGAAT CCCTGGCTCT 1600
 1601 ACCAGCTGTG AATCCGCTGG TGACGCCGCA GAAGGGAAGC  ACCCTGGACA ACAGCCTGCA CAAGGACCTG CTGGGCGCCA 1680
 1681 TCTCCGGCAT TGCTTCTCCA TATACAACCT CAACTCCAAA  CATCAACAGT GTGAGAAATG CTGATTCGAG AGGATCTCTC 1760
 1761 ATAAGCACAG ATTCGGGTAA CAGCCTTCCA GAAAGGAATA  GTGAGAAGAG CAATTCCCTG GATAAGCACC AACAAAGTAG 1840
 1841 CACATTGGGA AATTCCGTGG TTCGCTGTGA TAAACTTGAC  CAGTCTGAGA TTAAGAGCCT ACTGATGTGT TTCCTCTACA 1920
 1921 TCTTAAAGAG CATGTCTGAT GATGCTTTGT TTACATATTG  GAACAAGGCT TCAACATCTG AACTTATGGA TTTTTTTACA 2000
 2001 ATATCTGAAG TCTGCCTGCA CCAGTTCCAG TACATGGGGA  AGCGATACAT AGCCAGGAAC CAGGAGGGGT TGGGACCCAT 2080
 2081 AGTTCATGAT CGAAAGTCTC AGACATTGCC TGTTTCCCGT  AACAGAACAG GAATGATGCA TGCCAGATTG CAGCAGCTGG 2160
 2161 GCAGCCTGGA TAACTCTCTC ACTTTTAACC ACAGCTATGG  CCACTCGGAC GCAGATGTTC TGCACCAGTC ATTACTTGAA 2240
 2241 GCCAACATTG CTACTGAGGT TTGCCTGACA GCTCTGGACA  CGCTTTCTCT ATTTACATTG GCGTTTAAGA ACCAGCTCCT 2320
 2321 GGCCGACCAT GGACATAATC CTCTCATGAA AAAAGTTTTT  GATGTCTACC TGTGTTTTCT TCAAAAACAT CAGTCTGAAA 2400
 2401 CGGCTTTAAA AAATGTCTTC ACTGCCTTAA GGTCCTTAAT  TTATAAGTTT CCCTCAACAT TCTATGAAGG GAGAGCGGAC 2480
 2481 ATGTGTGCGG CTCTGTGTTA CGAGATTCTC AAGTGCTGTA  ACTCCAAGCT GAGCTCCATC AGGACGGAGG CCTCCCAGCT 2560
 2561 GCTCTACTTC CTGATGAGGA ACAACTTTGA TTACACTGGA  AAGAAGTCCT TTGTCCGGAC ACATTTGCAA GTCATCATAT 2640
 2641 CTGTCAGCCA GCTGATAGCA GACGTTGTTG GCATTGGGGA  AACCAGATTC CAGCAGTCCC TGTCCATCAT CAACAACTGT 2720
 2721 GCCAACAGTG ACCGGCTTAT TAAGCACACC AGCTTCTCCT  CTGATGTGAA GGACTTAACC AAAAGGATAC GCACGGTGCT 2800
 2801 AATGGCCACC GCCCAGATGA AGGAGCATGA GAACGACCCA  GAGATGCTGG TGGACCTCCA GTACAGCCTG GCCAAATCCT 2880
 2881 ATGCCAGCAC GCCCGAGCTC AGGAAGACGT GGCTCGACAG  CATGGCCAGG ATCCATGTCA AAAATGGCGA TCTCTCAGAG 2960
 2961 GCAGCAATGT GCTATGTCCA CGTAACAGCC CTAGTGGCAG  AATATCTCAC ACGGAAAGGC GTGTTTAGAC AAGGATGCAC 3040
 3041 CGCCTTCAGG GTCATTACCC CAAACATCGA CGAGGAGGCC  TCCATGATGG AAGACGTGGG GATGCAGGAT GTCCATTTCA 3120
 3121 ACGAGGATGT GCTGATGGAG CTCCTTGAGC AGTGCGCAGA  TGGACTCTGG AAAGCCGAGC GCTACGAGCT CATCGCCGAC 3200
 3201 ATCTACAAAC TTATCATCCC CATTTATGAG AAGCGGAGGG  ATTTCTTTGA AGATGAAGAT GGAAAGGAGT ATATTTACAA 3280
 3281 GGAACCCAAA CTCACACCGC TGTCGGAAAT TTCTCAGAGA  CTCCTTAAAC TGTACTCGGA TAAATTTGGT TCTGAAAATG 3360
 3361 TCAAAATGAT ACAGGATTCT GGCAAGGTCA ACCCTAAGGA  TCTGGATTCT AAGTATGCAT ACATCCAGGT GACTCACGTC 3440
 3441 ATCCCCTTCT TTGACGAAAA AGAGTTGCAA GAAAGGAAAA  CAGAGTTTGA GAGATCCCAC AACATCCGCC GCTTCATGTT 3520
 3521 TGAGATGCCA TTTACGCAGA CCGGGAAGAG GCAGGGCGGG  GTGGAAGAGC AGTGCAAACG GCGCACCATC CTGACAGCCA 3600
 3601 TACACTGCTT CCCTTATGTG AAGAAGCGCA TCCCTGTCAT  GTACCAGCAC CACACTGACC TGAACCCCAT CGAGGTGGCC 3680
 3681 ATTGACGAGA TGAGTAAGAA GGTGGCGGAG CTCCGGCAGC  TGTGCTCCTC GGCCGAGGTG GACATGATCA AACTGCAGCT 3760
 3761 CAAACTCCAG GGCAGCGTGA GTGTTCAGGT CAATGCTGGC  CCACTAGCAT ATGCGCGAGC TTTCTTAGAT GATACAAACA 3840
 3841 CAAAGCGATA TCCTGACAAT AAAGTGAAGC TGCTTAAGGA  AGTTTTCAGG CAATTTGTGG AAGCTTGCGG TCAAGCCTTA 3920
 3921 GCGGTAAACG AACGTCTGAT TAAAGAAGAC CAGCTCGAGT  ATCAGGAAGA AATGAAAGCC AACTACAGGG AAATGGCGAA 4000
 4001 GGAGCTTTCT GAAATCATGC ATGAGCAGAT CTGCCCCCTG  GAGGAGAAGA CGAGCGTCTT ACCGAATTCC CTTCACATCT 4080
 4081 TCAACGCCAT CAGTGGGACT CCAACAAGCA CAATGGTTCA  CGGGATGACC AGCTCGTCTT CGGTCGTGTG ATTACATCTC 4160
 4161 ATGGCCCGTG TGTGGGGACT TGCTTTGTCA TTTGCAAACT  CAGGATGCTT TCCAAAGCCA ATCACTGGGG AGACCGAGCA 4240
 4241 CAGGGAGGAC CAAGGGGAAG GGGAGAGAAA GGAAATAAAG  AACAACGTTA TTTCTTAACA GACTTTCTAT AGGAGTTGTA 4320
 4321 AGAAGGTGCA CATATTTTTT TAAATCTCAC TGGCAATATT  CAAAGTTTTC ATTGTGTCTT AACAAAGGTG TGGTAGACAC 4400
 4401 TCTTGAGCTG GACTTAGATT TTATTCTTCC TTGCAGAGTA  GTGTTAGAAT AGATGGCCTA CAGAAAAAAA AGGTTCTGGG 4480
 4481 ATCTACATGG CAGGGAGGGC TGCACTGACA TTGATGCCTG  GGGGACCTTT TGCCTCGACT CGTGCCGGAA ATCTGATCGT 4560
 4561 AATCAGGGTA CAGAACTTAC TAGTTTTGTC TAGGAGTATG  TTGTATGACT AGGATTTGTG CTATTATCTC ATTCAACAAC 4640
 4641 ATAGAGCAAG AATAGTGAGC TAACTGAGCT AGACACTCAA  TTAATCCGCT ACTGGCTTCA AGTCAGAACT TTGTCATTAA 4720
 4721 TCATCGACTC CGGGACGGTC ATATATGTAT TACATTTCTA  CATTTTTAAT ACTCACATGG GCTTATGCAT TAAGTTTAAT 4800
 4801 TGTGATAAAT TTGTGCTGGT CCAGTATATG CAATACACTT  TAATGGTTTA TTCTTGTCAT AAAAATGTGC AATATGGAGA 4880
 4881 TGTATACAAG TCTTTACT                                                                    4898
      |   10      |   20      |   30      |   40      |   50      |   60      |   70      |   80
```

FIG.10H

```
      |  10      |  20      |  30      |  40      |  50      |  60      |  70      |  80
    1 MEGHVMIAFL PTILNQLFRV LTRATQEEVA VNVTRVIIHV VAQCHEEGLE SHLRSYVKYA YKAEPYVASE YKTVHEELTK 80
   81 SMTTILKPSA DFLTSNKLLR YSWFFFDVLI KSMAQHLIEN SKVKLLRNQR FPASYHHAAE TVVNMLMPHI TQKFGDNPEA 160
  161 SKNANHSLAV FIKRCFTFMD RGFVFKQINN YISCFAPGDP KTLFEYKFEF LRVVCNHEHY IPLNLPMPFG KGRIQRYQDL 240
  241 QLDYSLTDEF CRNHFLVGLL LREVGTALQE FREVRLIAIS VLKNLLIKHS FDDRYASRSH QARIATLYLP LFGLLIENVQ 320
  321 RINVRDVSPF PVNAGMTVKD ESLALPAVNP LVTPQKGSTL DNSLHKDLLG AISGIASPYT TSTPNINSVR NADSRGSLIS 400
  401 TDSGNSLPER NSEKSNSLDK HQQSSTLGNS VVRCDKLDQS EIKSLLMCFL YILKSMSDDA LFTYWNKAST SELMDFFTIS 480
  481 EVCLHQFQYM GKRYIARNQE GLGPIVHDRK SQTLPVSRNR TGMMHARLQQ LGSLDNSLTF NHSYGHSDAD VLHQSLLEAN 560
  561 IATEVCLTAL DTLSLFTLAF KNQLLADHGH NPLMKKVFDV YLCFLQKHQS ETALKNVFTA LRSLIYKFPS TFYEGRADMC 640
  641 AALCYEILKC CNSKLSSIRT EASQLLYFLM RNNFDYTGKK SFVRTHLQVI ISVSQLIADV VGIGETRFQQ SLSIINNCAN 720
  721 SDRLIKHTSF SSDVKDLTKR IRTVLMATAQ MKEHENDPEM LVDLQYSLAK SYASTPELRK TWLDSMARIH VKNGDLSEAA 800
  801 MCYVHVTALV AEYLTRKGVF RQGCTAFRVI TPNIDEEASM MEDVGMQDVH FNEDVLMELL EQCADGLWKA ERYELIADIY 880
  881 KLIIPIYEKR RDFFEDEDGK EYIYKEPKLT PLSEISQRLL KLYSDKFGSE NVKMIQDSGK VNPKDLDSKY AYIQVTHVIP 960
  961 FFDEKELQER KTEFERSHNI RRFMFEMPFT QTGKRQGGVE EQCKRRTILT AIHCFPYVKK RIPVMYQHHT DLNPIEVAID 1040
 1041 EMSKKVAELR QLCSSAEVDM IKLQLKLQGS VSVQVNAGPL AYARAFLDDT NTKRYPDNKV KLLKEVFRQF VEACGQALAV 1120
 1121 NERLIKEDQL EYQEEMKANY REMAKELSEI MHEQICPLEE KTSVLPNSLH IFNAISGTPT STMVHGMTSS SSVV       1195
      |  10      |  20      |  30      |  40      |  50      |  60      |  70      |  80
```

FIG. 10H (cont.)

Exon 1A (-182 to -102)
GCAGGGGAAAAACCTGGCCCCATGATTCACTTACTTCCCACCGGATCTCTCCCATGACACGTGAGGATTA
TTACAATTTAA -102

Exon 1B (-219 to -102)
TTATCCCTTTACTACTTGCGAAGTGAGTTCGGTAGATGGGAGTGGAGAAGAGAACCTTAGAATCATTGTTTAGTCTTCAT
CTTTCACAGCTCAGGCTGAAGGCCTTTCCTTGCTGAGA -102

Exon 1C (-143 to -102)
GCGGCAGAGCGTGTCTGAGGTGGTGCGCGGCTCCGTGCTCCT -102

Exon2 and the rest of human CLASP2 cDNA

-101 -79
GGCAAAGCCAAAGCTAATTGAGC

-78 -1
AAGCTAATTGAGCCACTCGACTATGAAAATGTCATCGTCCAGAAGAAGACTCAGATCCTGAACGACTGTTTACGGGAG

1/1 31/11
ATG CTG CTC TTC CCT TAC GAT GAC TTT CAG ACG GCC ATC CTG AGA CGA CAG GGT CGA TAC
Met leu leu phe pro tyr asp asp phe gln thr ala ile leu arg arg gln gly arg tyr

61/21 91/31
ATA TGC TCA ACA GTG CCT GCG AAG GCG GAA GAG GAA GCA CAG AGC TTG TTT GTT ACA GAG
ile cys ser thr val pro ala lys ala glu glu glu ala gln ser leu phe val thr glu

121/41 151/51
TGC ATC AAA ACC TAT AAC TCT GAC TGG CAT CTT GTG AAC TAT AAA TAT GAA GAT TAC TCA
cys ile lys thr tyr asn ser asp trp his leu val asn tyr lys tyr glu asp tyr ser

181/61 211/71
GGA GAG TTT CGA CAG CTT CCG AAC AAA GTG GTC AAG TTG GAT AAA CTT CCA GTT CAT GTC
gly glu phe arg gln leu pro asn lys val val lys leu asp lys leu pro val his val

241/81 271/91
TAT GAA GTT GAC GAG GAG GTC GAC AAA GAT GAG GAT GCT GCC TCC CTT GGT TCC CAG AAG
tyr glu val asp glu glu val asp lys asp glu asp ala ala ser leu gly ser gln lys

301/101 331/111
GGT GGG ATC ACC AAG CAT GGC TGG CTG TAC AAA GGC AAC ATG AAC AGT GCC ATC AGC GTG
gly gly ile thr lys his gly trp leu tyr lys gly asn met asn ser ala ile ser val

361/121 391/131
ACC ATG AGG TCA TTT AAG AGA CGA TTT TTC CAC CTG ATT CAA CTT GGC GAT GGA TCC TAT
thr met arg ser phe lys arg arg phe phe his leu ile gln leu gly asp gly ser tyr

421/141 451/151
AAT TTG AAT TTT TAT AAA GAT GAA AAG ATC TCC AAA GAA CCA AAA GGA TCA ATA TTT CTG
asn leu asn phe tyr lys asp glu lys ile ser lys glu pro lys gly ser ile phe leu

481/161 511/171
GAT TCC TGT ATG GGT GTC GTT CAG AAC AAC AAA GTC AGG CGT TTT GCT TTT GAG CTC AAG
asp ser cys met gly val val gln asn asn lys val arg arg phe ala phe glu leu lys

FIG. 11A
1 of 8

```
541/181                                                 571/191
ATG CAG GAC AAA AGT AGT TAT CTC TTG GCA GCA GAC AGT GAA GTG GAA ATG GAA GAA TGG
met gln asp lys ser ser tyr leu leu ala ala asp ser glu val glu met glu glu trp

601/201                                                 631/211
ATC ACA ATT CTA AAT AAG ATC CTC CAG CTC AAC TTT GAA GCT GCA ATG CAA GAA AAG CGA
ile thr ile leu asn lys ile leu gln leu asn phe glu ala ala met gln glu lys arg

661/221                                                 691/231
AAT GGC GAC TCT CAC GAA GAT GAT GAA CAA AGC AAA TTG GAA GGT TCT GGT TCC GGT TTA
asn gly asp ser his glu asp asp glu gln ser lys leu glu gly ser gly ser gly leu

721/241                                                 751/251
GAT AGC TAC CTG CCG GAA CTT GCC AAG AGT GCA AGA GAA GCA GAA ATC AAA CTA AAA AGT
asp ser tyr leu pro glu leu ala lys ser ala arg glu ala glu ile lys leu lys ser

781/261                                                 811/271
GAA AGC AGA GTC AAA CTT TTT TAT TTG GAC CCA GAT GCC CAG AAG CTT GAC TTC TCA TCA
glu ser arg val lys leu phe tyr leu asp pro asp ala gln lys leu asp phe ser ser

841/281                                                 871/291
GCT GAG CCA GAA GTG AAG TCA TTT GAA GAG AAG TTT GGA AAA AGG ATC CTT GTC AAG TGC
ala glu pro glu val lys ser phe glu glu lys phe gly lys arg ile leu val lys cys

901/301                                                 931/311
AAT GAT TTA TCT TTC AAT TTG CAA TGC TGT GTT GCC GAA AAT GAA GAA GGA CCC ACT ACA
asn asp leu ser phe asn leu gln cys cys val ala glu asn glu glu gly pro thr thr

961/321                                                 991/331
AAT GTT GAA CCT TTC TTT GTT ACT CTA TCC CTG TTT GAC ATA AAA TAC AAC CGG AAG ATT
asn val glu pro phe phe val thr leu ser leu phe asp ile lys tyr asn arg lys ile

1021/341                                                1051/351
TCT GCC GAT TTC CAC GTA GAC CTG AAC CAT TTC TCA GTG AGG CAA ATG CTC GCC ACC ACG
ser ala asp phe his val asp leu asn his phe ser val arg gln met leu ala thr thr

1081/361                                                1111/371
TCC CCG GCG CTG ATG AAT GGC AGT GGG CAG AGC CCA TCT GTC CTC AAG GGC ATC CTT CAT
ser pro ala leu met asn gly ser gly gln ser pro ser val leu lys gly ile leu his

1141/381                                                1171/391
GAA GCC GCC ATG CAG TAT CCG AAG CAG GGA ATA TTT TCA GTC ACT TGT CCT CAT CCA GAT
glu ala ala met gln tyr pro lys gln gly ile phe ser val thr cys pro his pro asp

1201/401                                                1231/411
ATA TTT CTT GTG GCC AGA ATT GAA AAA GTC CTT CAG GGG AGC ATC ACA CAT TGC GCT GAG
ile phe leu val ala arg ile glu lys val leu gln gly ser ile thr his cys ala glu

1261/421                                                1291/431
CCA TAT ATG AAA AGT TCA GAC TCT TCT AAG GTG GCC CAG AAG GTG CTG AAG AAT GCC AAG
pro tyr met lys ser ser asp ser ser lys val ala gln lys val leu lys asn ala lys

1321/441                                                1351/451
CAG GCA TGC CAA AGA CTA GGA CAG TAT AGA ATG CCA TTT GCT TGG GCA GCA AGG ACA TTG
gln ala cys gln arg leu gly gln tyr arg met pro phe ala trp ala ala arg thr leu
```

FIG. 11A
2 of 8

```
1381/461                                                    1411/471
TTT AAG GAT GCA TCT GGA AAT CTT GAC AAA AAT GCC AGA TTT TCT GCC ATC TAC AGG CAA
phe lys asp ala ser gly asn leu asp lys asn ala arg phe ser ala ile tyr arg gln

1441/481                                                    1471/491
GAC AGC AAT AAG CTA TCC AAT GAT GAC ATG CTC AAG TTA CTT GCA GAC TTT CGG AAA CCT
asp ser asn lys leu ser asn asp asp met leu lys leu leu ala asp phe arg lys pro

1501/501                                                    1531/511
GAG AAG ATG GCT AAG CTC CCA GTG ATT TTA GGC AAT CTA GAC ATT ACA ATT GAT AAT GTT
glu lys met ala lys leu pro val ile leu gly asn leu asp ile thr ile asp asn val

1561/521                                                    1591/531
TCC TCA GAC TTC CCT AAT TAT GTT AAT TCA TCA TAC ATT CCC ACA AAA CAA TTT GAA ACC
ser ser asp phe pro asn tyr val asn ser ser tyr ile pro thr lys gln phe glu thr

1621/541                                                    1651/551
TGC AGT AAA ACT CCC ATC ACG TTT GAA GTG GAG GAA TTT GTG CCC TGC ATA CCA AAA CAC
cys ser lys thr pro ile thr phe glu val glu glu phe val pro cys ile pro lys his

1681/561                                                    1711/571
ACT CAG CCT TAC ACC ATC TAC ACC AAT CAC CTT TAC GTT TAT CCT AAG TAC TTG AAA TAC
thr gln pro tyr thr ile tyr thr asn his leu tyr val tyr pro lys tyr leu lys tyr

1741/581                                                    1771/591
GAC AGT CAG AAG TCT TTT GCC AAG GCT AGA AAT ATT GCG ATT TGC ATT GAA TTC AAA GAT
asp ser gln lys ser phe ala lys ala arg asn ile ala ile cys ile glu phe lys asp

1801/601                                                    1831/611
TCA GAT GAG GAA GAC TCT CAG CCC CTT AAG TGC ATT TAT GGC AGA CCT GGT GGG CCA GTT
ser asp glu glu asp ser gln pro leu lys cys ile tyr gly arg pro gly gly pro val

1861/621                                                    1891/631
TTC ACA AGA AGC GCC TTT GCT GCA GTT TTA CAC CAT CAC CAA AAC CCA GAA TTT TAT GAT
phe thr arg ser ala phe ala ala val leu his his his gln asn pro glu phe tyr asp

1921/641                                                    1951/651
GAG ATT AAA ATA GAG TTG CCC ACT CAG CTG CAT GAA AAG CAC CAC CTG TTG CTC ACA TTC
glu ile lys ile glu leu pro thr gln leu his glu lys his his leu leu leu thr phe

1981/661                                                    2011/671
TTC CAT GTC AGC TGT GAC AAC TCA AGT AAA GGA AGC ACG AAG AAG AGG GAT GTC GTT GAA
phe his val ser cys asp asn ser ser lys gly ser thr lys lys arg asp val val glu

2041/681                                                    2071/691
ACC CAA GTT GGC TAC TCC TGG CTT CCC CTC CTG AAA GAC GGA AGG GTG GTG ACA AGC GAG
thr gln val gly tyr ser trp leu pro leu leu lys asp gly arg val val thr ser glu

2101/701                                                    2131/711
CAG CAC ATC CCG GTC TCG GCG TAC CTT CCT TCG GGC CAT CTT GGC TAC CAA GAG CTT GGG
gln his ile pro val ser ala tyr leu pro ser gly his leu gly tyr gln glu leu gly

2161/721                                                    2191/731
ATG GGC AGG CAT TAT GGT CCG GAA ATT AAA TGG GTA GAT GGA GGC AAG CCA CTG CTG AAA
met gly arg his tyr gly pro glu ile lys trp val asp gly gly lys pro leu leu lys
```

FIG. 11A
3 of 8

```
2221/741                                                    2251/751
ATT TCC ACT CAT CTG GTT TCT ACA GTG TAT ACT CAG GAT CAG CAT TTA CAT AAT TTT TTC
ile ser thr his leu val ser thr val tyr thr gln asp gln his leu his asn phe phe

2281/761                                                    2311/771
CAG TAC TGT CAG AAA ACC GAA TCT GGA GCC CAA GCC TTA GGA AAC GAA CTT GTA AAG TAC
gln tyr cys gln lys thr glu ser gly ala gln ala leu gly asn glu leu val lys tyr

2341/781                                                    2371/791
CTT AAG AGT CTG CAT GCG ATG GAA GGC CAC GTG ATG ATC GCC TTC TTG CCC ACT ATC CTA
leu lys ser leu his ala met glu gly his val met ile ala phe leu pro thr ile leu

2401/801                                                    2431/811
AAC CAG CTG TTC CGA GTC CTC ACC AGA GCC ACA CAG GAA GAA GTC GCG GTT AAC GTG ACT
asn gln leu phe arg val leu thr arg ala thr gln glu glu val ala val asn val thr

2461/821                                                    2491/831
CGG GTC ATT ATT CAT GTG GTT GCC CAG TGC CAT GAG GAA GGA TTG GAG AGC CAC TTG AGG
arg val ile ile his val val ala gln cys his glu glu gly leu glu ser his leu arg

2521/841                                                    2551/851
TCA TAT GTT AAG TAC GCG TAT AAG GCT GAG CCA TAT GTT GCC TCT GAA TAC AAG ACA GTG
ser tyr val lys tyr ala tyr lys ala glu pro tyr val ala ser glu tyr lys thr val

2581/861                                                    2611/871
CAT GAA GAA CTG ACC AAA TCC ATG ACC ACG ATT CTC AAG CCT TCT GCC GAT TTC CTC ACC
his glu glu leu thr lys ser met thr thr ile leu lys pro ser ala asp phe leu thr

2641/881                                                    2671/891
AGC AAC AAA CTA CTG AGG TAC TCA TGG TTT TTC TTT GAT GTA CTG ATC AAA TCT ATG GCT
ser asn lys leu leu arg tyr ser trp phe phe phe asp val leu ile lys ser met ala

2701/901                                                    2731/911
CAG CAT TTG ATA GAG AAC TCC AAA GTT AAG TTG CTG CGA AAC CAG AGA TTT CCT GCA TCC
gln his leu ile glu asn ser lys val lys leu leu arg asn gln arg phe pro ala ser

2761/921                                                    2791/931
TAT CAT CAT GCA GCG GAA ACC GTT GTA AAT ATG CTG ATG CCA CAC ATC ACT CAG AAG TTT
tyr his his ala ala glu thr val val asn met leu met pro his ile thr gln lys phe

2821/941                                                    2851/951
GGA GAT AAT CCA GAG GCA TCT AAG AAC GCG AAT CAT AGC CTT GCT GTC TTC ATC AAG AGA
gly asp asn pro glu ala ser lys asn ala asn his ser leu ala val phe ile lys arg

2881/961                                                    2911/971
TGT TTC ACC TTC ATG GAC AGG GGC TTT GTC TTC AAG CAG ATC AAC AAC TAC ATT AGC TGT
cys phe thr phe met asp arg gly phe val phe lys gln ile asn asn tyr ile ser cys

2941/981                                                    2971/991
TTT GCT CCT GGA GAC CCA AAG ACC CTC TTT GAA TAC AAG TTT GAA TTT CTC CGT GTA GTG
phe ala pro gly asp pro lys thr leu phe glu tyr lys phe glu phe leu arg val val

3001/1001                                                   3031/1011
TGC AAC CAT GAA CAT TAT ATT CCG TTG AAC TTA CCA ATG CCA TTT GGA AAA GGC AGG ATT
cys asn his glu his tyr ile pro leu asn leu pro met pro phe gly lys gly arg ile
```

FIG. 11A
4 of 8

```
3061/1021                                          3091/1031
CAA AGA TAC CAA GAC CTC CAG CTT GAC TAC TCA TTA ACA GAT GAG TTC TGC AGA AAC CAC
gln arg tyr gln asp leu gln leu asp tyr ser leu thr asp glu phe cys arg asn his

3121/1041                                          3151/1051
TTC TTG GTG GGA CTG TTA CTG AGG GAG GTG GGG ACA GCC CTC CAG GAG TTC CGG GAG GTC
phe leu val gly leu leu leu arg glu val gly thr ala leu gln glu phe arg glu val

3181/1061                                          3211/1071
CGT CTG ATC GCC ATC AGT GTG CTC AAG AAC CTG CTG ATA AAG CAT TCT TTT GAT GAC AGA
arg leu ile ala ile ser val leu lys asn leu leu ile lys his ser phe asp asp arg

3241/1081                                          3271/1091
TAT GCT TCA AGG AGC CAT CAG GCA AGG ATA GCC ACC CTC TAC CTG CCT CTG TTT GGT CTG
tyr ala ser arg ser his gln ala arg ile ala thr leu tyr leu pro leu phe gly leu

3301/1101                                          3331/1111
CTG ATT GAA AAC GTC CAG CGG ATC AAT GTG AGG GAT GTG TCA CCC TTC CCT GTG AAC GCG
leu ile glu asn val gln arg ile asn val arg asp val ser pro phe pro val asn ala

3361/1121                                          3391/1131
GGC ATG ACC GTG AAG GAT GAA TCC CTG GCT CTA CCA GCT GTG AAT CCG CTG GTG ACG CCG
gly met thr val lys asp glu ser leu ala leu pro ala val asn pro leu val thr pro

3421/1141                                          3451/1151
CAG AAG GGA AGC ACC CTG GAC AAC AGC CTG CAC AAG GAC CTG CTG GGC GCC ATC TCC GGC
gln lys gly ser thr leu asp asn ser leu his lys asp leu leu gly ala ile ser gly

3481/1161                                          3511/1171
ATT GCT TCT CCA TAT ACA ACC TCA ACT CCA AAC ATC AAC AGT GTG AGA AAT GCT GAT TCG
ile ala ser pro tyr thr thr ser thr pro asn ile asn ser val arg asn ala asp ser

3541/1181                                          3571/1191
AGA GGA TCT CTC ATA AGC ACA GAT TCG GGT AAC AGC CTT CCA GAA AGG AAT AGT GAG AAG
arg gly ser leu ile ser thr asp ser gly asn ser leu pro glu arg asn ser glu lys

3601/1201                                          3631/1211
AGC AAT TCC CTG GAT AAG CAC CAA CAA AGT AGC ACA TTG GGA AAT TCC GTG GTT CGC TGT
ser asn ser leu asp lys his gln gln ser ser thr leu gly asn ser val val arg cys

3661/1221                                          3691/1231
GAT AAA CTT GAC CAG TCT GAG ATT AAG AGC CTA CTG ATG TGT TTC CTC TAC ATC TTA AAG
asp lys leu asp gln ser glu ile lys ser leu leu met cys phe leu tyr ile leu lys

3721/1241                                          3751/1251
AGC ATG TCT GAT GAT GCT TTG TTT ACA TAT TGG AAC AAG GCT TCA ACA TCT GAA CTT ATG
ser met ser asp asp ala leu phe thr tyr trp asn lys ala ser thr ser glu leu met

3781/1261                                          3811/1271
GAT TTT TTT ACA ATA TCT GAA GTC TGC CTG CAC CAG TTC CAG TAC ATG GGG AAG CGA TAC
asp phe phe thr ile ser glu val cys leu his gln phe gln tyr met gly lys arg tyr

3841/1281                                          3871/1291
ATA GCC AGG AAC CAG GAG GGG TTG GGA CCC ATA GTT CAT GAT CGA AAG TCT CAG ACA TTG
ile ala arg asn gln glu gly leu gly pro ile val his asp arg lys ser gln thr leu
```

FIG. 11A
5 of 8

```
3901/1301                                                    3931/1311
CCT GTT TCC CGT AAC AGA ACA GGA ATG ATG CAT GCC AGA TTG CAG CAG CTG GGC AGC CTG
pro val ser arg asn arg thr gly met met his ala arg leu gln gln leu gly ser leu

3961/1321                                                    3991/1331
GAT AAC TCT CTC ACT TTT AAC CAC AGC TAT GGC CAC TCG GAC GCA GAT GTT CTG CAC CAG
asp asn ser leu thr phe asn his ser tyr gly his ser asp ala asp val leu his gln

4021/1341                                                    4051/1351
TCA TTA CTT GAA GCC AAC ATT GCT ACT GAG GTT TGC CTG ACA GCT CTG GAC ACG CTT TCT
ser leu leu glu ala asn ile ala thr glu val cys leu thr ala leu asp thr leu ser

4081/1361                                                    4111/1371
CTA TTT ACA TTG GCG TTT AAG AAC CAG CTC CTG GCC GAC CAT GGA CAT AAT CCT CTC ATG
leu phe thr leu ala phe lys asn gln leu leu ala asp his gly his asn pro leu met

4141/1381                                                    4171/1391
AAA AAA GTT TTT GAT GTC TAC CTG TGT TTT CTT CAA AAA CAT CAG TCT GAA ACG GCT TTA
lys lys val phe asp val tyr leu cys phe leu gln lys his gln ser glu thr ala leu

4201/1401                                                    4231/1411
AAA AAT GTC TTC ACT GCC TTA AGG TCC TTA ATT TAT AAG TTT CCC TCA ACA TTC TAT GAA
lys asn val phe thr ala leu arg ser leu ile tyr lys phe pro ser thr phe tyr glu

4261/1421                                                    4291/1431
GGG AGA GCG GAC ATG TGT GCG GCT CTG TGT TAC GAG ATT CTC AAG TGC TGT AAC TCC AAG
gly arg ala asp met cys ala ala leu cys tyr glu ile leu lys cys cys asn ser lys

4321/1441                                                    4351/1451
CTG AGC TCC ATC AGG ACG GAG GCC TCC CAG CTG CTC TAC TTC CTG ATG AGG AAC AAC TTT
leu ser ser ile arg thr glu ala ser gln leu leu tyr phe leu met arg asn asn phe

4381/1461                                                    4411/1471
GAT TAC ACT GGA AAG AAG TCC TTT GTC CGG ACA CAT TTG CAA GTC ATC ATA TCT GTC AGC
asp tyr thr gly lys lys ser phe val arg thr his leu gln val ile ile ser val ser

4441/1481                                                    4471/1491
CAG CTG ATA GCA GAC GTT GTT GGC ATT GGG GAA ACC AGA TTC CAG CAG TCC CTG TCC ATC
gln leu ile ala asp val val gly ile gly glu thr arg phe gln gln ser leu ser ile

4501/1501                                                    4531/1511
ATC AAC AAC TGT GCC AAC AGT GAC CGG CTT ATT AAG CAC ACC AGC TTC TCC TCT GAT GTG
ile asn asn cys ala asn ser asp arg leu ile lys his thr ser phe ser ser asp val

4561/1521                                                    4591/1531
AAG GAC TTA ACC AAA AGG ATA CGC ACG GTG CTA ATG GCC ACC GCC CAG ATG AAG GAG CAT
lys asp leu thr lys arg ile arg thr val leu met ala thr ala gln met lys glu his

4621/1541                                                    4651/1551
GAG AAC GAC CCA GAG ATG CTG GTG GAC CTC CAG TAC AGC CTG GCC AAA TCC TAT GCC AGC
glu asn asp pro glu met leu val asp leu gln tyr ser leu ala lys ser tyr ala ser

4681/1561                                                    4711/1571
ACG CCC GAG CTC AGG AAG ACG TGG CTC GAC AGC ATG GCC AGG ATC CAT GTC AAA AAT GGC
thr pro glu leu arg lys thr trp leu asp ser met ala arg ile his val lys asn gly
```

FIG. 11A
6 of 8

```
4741/1581                                                  4771/1591
GAT CTC TCA GAG GCA GCA ATG TGC TAT GTC CAC GTA ACA GCC CTA GTG GCA GAA TAT CTC
asp leu ser glu ala ala met cys tyr val his val thr ala leu val ala glu tyr leu

4801/1601                                                  4831/1611
ACA CGG AAA GGC GTG TTT AGA CAA GGA TGC ACC GCC TTC AGG GTC ATT ACC CCA AAC ATC
thr arg lys gly val phe arg gln gly cys thr ala phe arg val ile thr pro asn ile

4861/1621                                                  4891/1631
GAC GAG GAG GCC TCC ATG ATG GAA GAC GTG GGG ATG CAG GAT GTC CAT TTC AAC GAG GAT
asp glu glu ala ser met met glu asp val gly met gln asp val his phe asn glu asp

4921/1641                                                  4951/1651
GTG CTG ATG GAG CTC CTT GAG CAG TGC GCA GAT GGA CTC TGG AAA GCC GAG CGC TAC GAG
val leu met glu leu leu glu gln cys ala asp gly leu trp lys ala glu arg tyr glu

4981/1661                                                  5011/1671
CTC ATC GCC GAC ATC TAC AAA CTT ATC ATC CCC ATT TAT GAG AAG CGG AGG GAT TTC TTT
leu ile ala asp ile tyr lys leu ile ile pro ile tyr glu lys arg arg asp phe phe

5041/1681                                                  5071/1691
GAA GAT GAA GAT GGA AAG GAG TAT ATT TAC AAG GAA CCC AAA CTC ACA CCG CTG TCG GAA
glu asp glu asp gly lys glu tyr ile tyr lys glu pro lys leu thr pro leu ser glu

5101/1701                                                  5131/1711
ATT TCT CAG AGA CTC CTT AAA CTG TAC TCG GAT AAA TTT GGT TCT GAA AAT GTC AAA ATG
ile ser gln arg leu leu lys leu tyr ser asp lys phe gly ser glu asn val lys met

5161/1721                                                  5191/1731
ATA CAG GAT TCT GGC AAG GTC AAC CCT AAG GAT CTG GAT TCT AAG TAT GCA TAC ATC CAG
ile gln asp ser gly lys val asn pro lys asp leu asp ser lys tyr ala tyr ile gln

5221/1741                                                  5251/1751
GTG ACT CAC GTC ATC CCC TTC TTT GAC GAA AAA GAG TTG CAA GAA AGG AAA ACA GAG TTT
val thr his val ile pro phe phe asp glu lys glu leu gln glu arg lys thr glu phe

5281/1761                                                  5311/1771
GAG AGA TCC CAC AAC ATC CGC CGC TTC ATG TTT GAG ATG CCA TTT ACG CAG ACC GGG AAG
glu arg ser his asn ile arg arg phe met phe glu met pro phe thr gln thr gly lys

5341/1781                                                  5371/1791
AGG CAG GGC GGG GTG GAA GAG CAG TGC AAA CGG CGC ACC ATC CTG ACA GCC ATA CAC TGC
arg gln gly gly val glu glu gln cys lys arg arg thr ile leu thr ala ile his cys

5401/1801                                                  5431/1811
TTC CCT TAT GTG AAG AAG CGC ATC CCT GTC ATG TAC CAG CAC CAC ACT GAC CTG AAC CCC
phe pro tyr val lys lys arg ile pro val met tyr gln his his thr asp leu asn pro

5461/1821                                                  5491/1831
ATC GAG GTG GCC ATT GAC GAG ATG AGT AAG AAG GTG GCG GAG CTC CGG CAG CTG TGC TCC
ile glu val ala ile asp glu met ser lys lys val ala glu leu arg gln leu cys ser

5521/1841                                                  5551/1851
TCG GCC GAG GTG GAC ATG ATC AAA CTG CAG CTC AAA CTC CAG GGC AGC GTG AGT GTT CAG
ser ala glu val asp met ile lys leu gln leu lys leu gln gly ser val ser val gln
```

FIG. 11A
7 of 8

```
5581/1861                                          5611/1871
GTC AAT GCT GGC CCA CTA GCA TAT GCG CGA GCT TTC TTA GAT GAT ACA AAC ACA AAG CGA
val asn ala gly pro leu ala tyr ala arg ala phe leu asp asp thr asn thr lys arg

5641/1881                                          5671/1891
TAT CCT GAC AAT AAA GTG AAG CTG CTT AAG GAA GTT TTC AGG CAA TTT GTG GAA GCT TGC
tyr pro asp asn lys val lys leu leu lys glu val phe arg gln phe val glu ala cys

5701/1901                                          5731/1911
GGT CAA GCC TTA GCG GTA AAC GAA CGT CTG ATT AAA GAA GAC CAG CTC GAG TAT CAG GAA
gly gln ala leu ala val asn glu arg leu ile lys glu asp gln leu glu tyr gln glu

5761/1921                                          5791/1931
GAA ATG AAA GCC AAC TAC AGG GAA ATG GCG AAG GAG CTT TCT GAA ATC ATG CAT GAG CAG
glu met lys ala asn tyr arg glu met ala lys glu leu ser glu ile met his glu gln

5821/1941                                          5851/1951
ATC TGC CCC CTG GAG GAG AAG ACG AGC GTC TTA CCG AAT TCC CTT CAC ATC TTC AAC GCC
ile cys pro leu glu glu lys thr ser val leu pro asn ser leu his ile phe asn ala

5881/1961                                          5911/1971
ATC AGT GGG ACT CCA ACA AGC ACA ATG GTT CAC GGG ATG ACC AGC TCG TCT TCG GTC GTG
ile ser gly thr pro thr ser thr met val his gly met thr ser ser ser ser val val

5941/1981                                          5971
TGA TTA CAT CTC ATG GCC CGT GTG TGG GGA CTT GCT TTG TCA TTT GCA AAC TCA GGA TGC
:TA

6001                                               6031
TTT CCA AAG CCA ATC ACT GGG GAG ACC GAG CAC AGG GAG GAC CAA GGG GAA GGG GAG AGA
6061                                               6091
AAG GAA ATA AAG AAC AAC GTT ATT TCT TAA CAG ACT TTC TAT AGG AGT TGT AAG AAG GTG
6121                                               6151
CAC ATA TTT TTT TAA ATC TCA CTG GCA ATA TTC AAA GTT TTC ATT GTG TCT TAA CAA AGG
6181                                               6211
TGT GGT AGA CAC TCT TGA GCT GGA CTT AGA TTT TAT TCT TCC TTG CAG AGT AGT GTT AGA
6241                                               6271
ATA GAT GGC CTA CAG AAA AAA AAG GTT CTG GGA TCT ACA TGG CAG GGA GGG CTG CAC TGA
6301                                               6331
CAT TGA TGC CTG GGG GAC CTT TTG CCT CGA CTC GTG CCG GAA ATC TGA TCG TAA TCA GGG
6361                                               6391
TAC AGA ACT TAC TAG TTT TGT CTA GGA GTA TGT TGT ATG ACT AGG ATT TGT GCT ATT ATC
6421                                               6451
TCA TTC AAC AAC ATA GAG CAA GAA TAG TGA GCT AAC TGA GCT AGA CAC TCA ATT AAT CCG
6481                                               6511
CTA CTG GCT TCA AGT CAG AAC TTT GTC ATT AAT CAT CGA CTC CGG GAC GGT CAT ATA TGT
6541                                               6571
ATT ACA TTT CTA CAT TTT TAA TAC TCA CAT GGG CTT ATG CAT TAA GTT TAA TTG TGA TAA
6601                                               6631
ATT TGT GCT GGT CCA GTA TAT GCA ATA CAC TTT AAT GGT TTA TTC TTG TCA TAA AAA TGT
6661
GCA ATA TGG AGA TGT ATA CAA GTC TTT ACT
```

FIG. 11A
8 of 8

A. Allelic variations: single nucleotide changes (polymorphism) between CLASP-2 cDNA isoforms

Isoform	Difference	Nucleotide(s)	Consequence
1	polymorphism	862	A to G change; mis-sense mutation
2	polymorphism		A to C change; mis-sense mutation changing codon from histidine to proline
3	polymorphism	2210	A to G change; mis-sense mutation changing codon from aspargine to glutamic acid
4	polymorphism	2225	C to T change; mis-sense mutation changing codon from histidine to tyrosine

B. Alternative splices

Isoform	Difference	Nucleotide(s)	Consequence
1	exon deletion	209-291	premature, in-frame stop codon leading to the production of a truncated, most likely soluble protein

These differences may be found separately or together in various combinations in the different human CLASP-2 isoforms

FIG. 11B

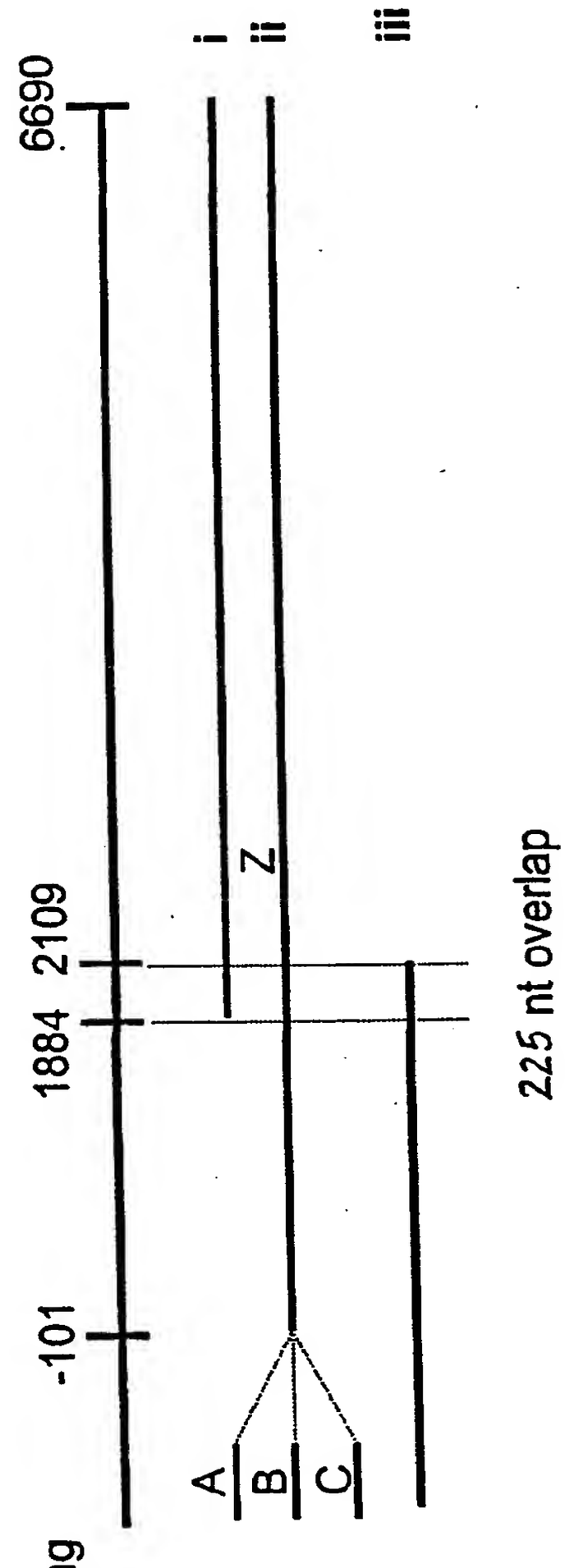

6690
i
ii
iii
-101
1884
2109
Z
A
B
C
225 nt overlap
human CLASP-2
nucleotide numbering
as in Figure 11A

FIG. 11C

1st exon (nucleotides 335 to 445)
TGTCTTGCTTATCTTTTCGCCCTCCAGGCAAAGCCAAAGCTAATTGAGCCACT
CGACTATGAAAATGTCATCGTCCAGAAGAAGACTCAGATCCTGAACGACTGT
TTACGGGAGATGCTGCTCTTCCCTTACGATGACTTTCAGGTAAGTAACGTTAT
GTTTCTATCCGTAGAACCACG

2nd exon (nucleotides 7101-7190)
TTACCCAAGGCTTTTCCTCCTGTTTTTGTTTCCAGACGGCCATCCTGAGACGA
CAGGGTCGATACATATGCTCAACAGTGCCTGCGAAGGCGGAAGAGGAAGCA
CAGAGCTTGTTTGTTACAGAGGTAAGGCTCTTTCCTGCATTAATTTACATTTT
GAAGTCATTTTCCCCTAACTGCCTCC

3rd exon (nucleotides 11439 to 11521)
TTTTCTATTTTTTAAAATCCCCCTTCAATAGTGCATCAAAACCTATAACTCTGAC
TGGCATCTTGTGAACTATAAATATGAAGATTACTCAGGAGAGTTTCGACAGC
TTCCGAAGTGAGTAAGCTATATTATACACATAGGGAAAAGTCTTT

4th exon (nucleotides 13987 to 14056)
CTAAAACAAATTTTCTTTGTTGTTTTTATAGCAAAGTGGTCAAGTTGGATAAA
CTTCCAGTTCATGTCTATGAAGTTGACGAGGAGGTCGACAAAGATGAGGTGG
GATACCTGCTTGCTGTTGCTTCTCTTTTCACTCTAGATTTAA

5th exon (nucleotides 15212 to 15307)
GGAGGTTGACTGCTGGTGTTTTCCTTCTCTCCTAGGATGCTGCCTCCCTTGGTT
CCCAGAAGGGTGGGATCACCAAGCATGGCTGGCTGTACAAAGGCAACATGA
ACAGTGCCATCAGCGTGACCATGAGGGTGAGGACGCACATCACTTTGCCCTC
CCCTCTCACAAGCCCTTTC

6th exon (nucleotides 16269 to 16404)
TGAAAGAATAGCTGTGTGTATATTTTTCTCTCAGTCATTTAAGAGACGATTTT
TCCACCTGATTCAACTTGGCGATGGATCCTATAATTTGAATTTTTATAAAGAT
GAAAAGATCTCCAAAGAACCAAAAGGATCAATATTTCTGGATTCCTGTATGG
GTGTCGTTCAGGTAAATATGAAAAGAGTTTTACCATTATGTTTTCTTA

7th exon (nucleotides 19459 to 19633)
AAGTATGTCTGTTTATCCTTTTTTCATTTCAGAACAACAAAGTCAGGCGTTTT
GCTTTTGAGCTCAAGATGCAGGACAAAAGTAGTTATCTCTTGGCAGCAGACA
GTGAAGTGGAAATGGAAGAATGGATCACAATTCTAAATAAGATCCTCCAGCT
CAACTTTGAAGCTGCAATGCAAGAAAAGCGAAATGGCGACTCTCACGAAGGT
AGATAGGCTTGGCTTCCCCCAGGCACATACACACTCT

8th exon (nucleotides 20567 to 20634)

FIG. 12A
1 of 2

ATTACAAGTGATTCCGATAATCTGTTTTGCCATTTTAGATGATGAACAAAGCA
AATTGGAAGGTTCTGGTTCCGGTTTAGATAGCTACCTGCCGGAACTTGCCAAG
GTAACATCGTCTTATATCTTCTGCTCTTCGTTGAATGC

9th exon (nucleotides 30257 to 30331)
GATTGTGTTAAATGTAATTTTCATGTATCTTGTTATCAGAGTGCAAGAGAAGC
AGAAATCAAACTAAAAAGTGAAAGCAGAGTCAAACTTTTTTATTTGGACCCA
GATGCCCAGGTAAGAACTATCTAAATGTTTAATATTTAAAACCAAAT

10th exon (nucleotides 31851 to 31991)
CATAACTTATTTATATGTTTACATTTTCTTTTAAAGAAGCTTGACTTCTCATCA
GCTGAGCCAGAAGTGAAGTCATTTGAAGAGAAGTTTGGAAAAAAGGATCCTTG
TCAAGTGCAATGATTTATCTTTCAATTTGCAATGCTGTGTTGCCGAAAATGAA
GAAGGACCCACTACAAATGTAATTTTTCATTTTAAAAATAAACATTAAAAAA
AAAATAGGCAG

11th exon (nucleotides 32472 to 32675)
CCATGGTGATCATTGGATTGTTTTGTTTTGTTCAGGTTGAACCTTTCTTTGTTA
CTCTATCCCTGTTTGACATAAAATACAACCGGAAGATTTCTGCCGATTTCCAC
GTAGACCTGAACCATTTCTCAGTGAGGCAAATGCTCGCCACCACGTCCCCGG
CGCTGATGAATGGCAGTGGGCAGAGCCCATCTGTCCTCAAGGGCATCCTTCA
TGAAGCCGCCATGCAGTATCCGAAGCAGGTGGGGAGTATGAGCCCAGCATTC
CCACTACTCAGACTCACTTTGCATGC

12th exon (nucleotides 33063 to 33185)
GAATTCTGCTTACTGAAGAAAATTGTTTGCCTCCTAGGGAATATTTTCAGTCA
CTTGTCCTCATCCAGATATATTTCTTGTGGCCAGAATTGAAAAAGTCCTTCAG
GGGAGCATCACACATTGCGCTGAGCCATATATGAAAAGTTCAGACTCTTCTA
AGGTATGAATGGCTTTTACGCTTTGGGGTGGTAAAAAGCAATCTGAA

13th exon (nucleotides 36702 to 36784)
CAGTATCTCATAGCTTTATTCTCATGTCTTCAAGGTGGCCCAGAAGGTGCTGA
AGAATGCCAAGCAGGCATGCCAAAGACTAGGACAGTATAGAATGCCATTTGC
TTGGGCAGCAAGGTAAGGAACACCTTTTATACCTTTTAAATCGATATAGATA
GGTGCATGG

14th partial exon (nucleotides 37353 to 37475)
GAAACCCAGTTTAGAAATGTTGCTTTGCCATTTCAGGACATTGTTTAAGGATG
CATCTGGAAATCTTGACAAAAATGCCAGATTTTCTGCCATCTACAGGCAAGA
CAGCAATAAGCTATCCAATGATGACATGCTCAAGTTACTTGCAGACTTTCGG
AA

FIG. 12A
2 of 2

```
1     TACCAAGGGCAACTCTGGCACACCCTAAAGTCTGGAAAGGGGACATAGCTAGTCAGGGATGACCCGAGAAATGACTGGAAGCTCCACCAGAA
93    TGCAGAGCTTCCTTTGTGCTTAAATAACTGAACAAGCATCACTCTGTGTAGCAGGACACCACCCAGCATTTTTTGTCCCTTTGGAAACAACT
185   CTTATTTCTGTTTCTTTGTGATACCAAAACTAGCATACTCTAATTGTAGAAAATACAAAACATAGAGTAGAACATACTAAGTTCTTTATCTT
277   AAGAAATGGCATTTGTGTATGAGAATGTCTTGCTTATCTTTTCGCCCTCCAGGCAAAGCCAAAGCTAATTGAGCCACTCGACTATGAAAATG
369   TCATCGTCCAGAAGAAGACTCAGATCCTGAACGACTGTTTACGGGAGATGCTGCTCTTCCCTTACGATGACTTTCAGGTAAGTAACGTTATG
461   TTTCTATCCGTAGAACCACGTGTTTGATCTTAACAAGCAGTATTTTTCTATGTATTGATTTATTGTTTGGTTAGTTAATTATTATTATTATT
553   ATTTTTTTTGAGACACAGTCTTGCTCTGTCACGCAGGCTTGAGTGCAGTGGTGCCATCTTGGCTCAACGGCAACCTCCGCCTCCTGGGTTCA
645   AGCATTTCTCCTGCCTCAGCCTCTCAAGTAACTGGGATTACAGGCGTGTGCCACCATGCCTGGCTAATTTTTGTCTTTGTATTAGAGACAGG
737   GTTTTGCCACGGTGGCCAGGGTCGTCCCAAACTCCTGGCCTTAAGTGATCTATCTGCCTTGGCCTCTCAAAATGTTGGGATTATAGGCATGA
829   GCCACTGTGCCCGGCCTAATTATGGTTTTTAAAAGATGAAAATAAGATGTTATTTAAGAAAGAAAAGTTATTTTATATTCTTCCAAGCATCC
921   TTCATGAGTTGATAATTTTTAATGGTATTATTTTTGCATATTAATTATAAGTATGCCAAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT
1013  GTGTATCTTGAATAAAAGTGCTATACTCTGTCTGGGCTGATTTAGTGGGCACAAGTGCCTTTCTGTCTTTTGAGATTTGTTTTGATTTAGAT
1105  TTTTTGGCCAAGTACCTTAGAATCTTAATGATGGGCTTTGCTGGGGAGCAGTGGGAGATTTCGTCATGCCTTTTTTTTAGGATGGCATTTGGG
1197  AGCCTGCCTTCAGAGGCCCTGCACTCTGTGTGGGCTCCCAGCAAAGCGTTCAAGGTTAGCCAAGAATGGCCTGAAGTTTACCTCTGTAGTGT
1289  AATGTGGGTGCTGTTCCTTGGAAGAAATGTGGAGGACTCAGCACAGCTCCTGCTGTGTGCCGCTCTTTCAGGCTATGGCCTGTGGTTAAGAG
1381  ACTAACAAGAAGCTGTGAGGCTGTTGAGGAATGAGAATGACATTCTTCCTCCAGGAAACCCGGTGGTGTTAAATGCCTTGCAGCGAGCCACC
1473  TTGGTCCATTTGGAGGTTCTGGTTACTTTCCTTGCTCTCCTGGGACCTGATCCTGGCACTTCTTTCTCCTCTCCTTTGATGCTCTTAGTTGG
1565  ACACTTCTCTACCTGATGCTGTTACCATTTAAGCCCTGTCTTTTGTGAATCGAGCTGCCTTTTTTTTTTTAAGCTTCACTGATTCTTTGTTGT
1657  TTGATTCCAAAAGTGTTACATCCATGTACAAAAAGATAAATGAGAGGGAAATATTGAAATAATTGACATGAAAAGCCTCCCCACGCCTTCTA
1749  ATCCCATCCCACAGAACATGACTTAACTGTATACAGCTCTGTGCATACTTGTCTTTAGAAACTTCCATGTTAATAGAAATTGTTAAATTACG
1841  ATCCTTGAAGGTTTTTTCCCACCAAATTTAAGCGACTCCAGCTTACAACAGAGGTGAGAATTTTCACAAATGTTCACTCTTTCTAACTTGTT
1933  AGAGATACCTGGGCCCCAAAATGATTATTCTTTAGCTCTGTCCTGCATAAAAGGAATGCCCATGGGAATGAAATTGACCATTCGTGTGGTGT
2025  TGCTACCAAAGTAACAGGTAAATGGGTTGAGGTCATGCCAAACAATACCATGCTTTGCATACTTCATTTCATGACTAAACTGCATGGGAACG
2117  GACTAATAAATGAGAACCTCTGAATGATGCCTTTTGCCTGTGATTTGGCAACAAATGAAAAGCAAAATCAAATGATTATAAATTGTACTGCA
2209  TGTTGACAAGATTTTCCTGTAGTGTTGTCTGAGGAAGCTAAAGGTTATCTCAAATTTCTCTCAACATGAAGTATGTGTTCTTCTTGGTATTA
2301  ATTAAAGTAACAACTTTTTTGAGTTTGCAACCTAGAATGAAAAATTCTATTTGTATGACTGAGATAAAATTGCTTAAGAAACAACCAAAGAA
2393  ACGAGATACAGTTAGTTGAGTGTCATCTTTATCCCAGGGAACAGGTATCTGGATGTTTAAGCAGTTGCAGAATCAGACAGTTTAAACTTTGA
2485  GAAAACTTCTGTGTCCCCTTTGCTTTTAACTACTCTGGTGATAGCAGGCACAAATATTCTAGGAAAGGCAAAGAACTCACTAGCATTTTGTT
2577  GGCTAAGGTGATGAGCAAATATTATTTTCTGTTTGGGGAGAAGTTTTCCTAGAGATTTAGGAGCTTGAATTGGAGCTTTAATCCTCATCACA
2669  GGAATTGTGATGGGCCCCAGTGAAGTTTGGGTACAATTATTTGTTTTCTTATAGACTCCCACTTTCTTATCAGGTAAAGCCATGTACTCTGT
2761  GCTTTCTTGTTAAATTGTCTCAGTGATGTTATTAACTGTCTAATTAGCTGGATGAGTGAAAGGTCTTAACAGTGCCACAGATTCTTTCTATC
2853  TGTGTTTTCTTAGGCAGAATAAGAGCAGAATTATTGTATTATTAGAGGCAGAGGGAACAAATTAGATTGGGGAAAGTGTTTTATTTCATATG
2945  GAAAAGTAATACCAAGTTGGTTAGGAAATGGCAGCAGCAAAACGCATGCTGAGGGGTGATTTACTGCACTTAAATAATTTAGCAGTATAAGT
3037  TAACTATTAAAATAATAGAACTTGGTGTCCATTTCTGCCAAATATATTTGAAATGACAATTTACTAAAATATAAGCATGGATAGTGGTGATG
3129  CTTGTGTACATTTTTCAAGTAGGCACATGTTGATCTTGAGCCTTTACTGGTCAGATCCTAAAGGCATCTACATGTTCTCTAAAAATGAGTTG
3221  TGTCAAGAAAAGATTTGCGGGTTGCATGTAGTTGCCTGAGGATGACAGAAGAGTAGTTACTACAACAGCAGCAAAGAAGAGAGACATGAAGT
3313  AAACGTGGATTTTTAAAAATCAAAAGAATAGGCCAGGCGCACTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCAGATC
3405  ACAAGGTCAGGAGTTTGAGACCAGCCTGGCCAATATGGTGAAACCCCATCTCTACCAGAAAATACAAAATTAGCCAGGCATGGTGGTGCATG
3497  CCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCTCTTGAAACCAGGAGACAGAGGTTGCAGTAAGCTGAGATCGTGCCACTGCA
3589  TTCCAGCCTGGGCGACAGAGTGAGACTCCATCTCAAAACAAACAAAAAATCAAAATAATAGTTCCCAGCCATCAGGTTATTGATGAAGTAGG
3681  CTGGGCACGGTGGCTCACACCTGTAATCCCAGCACATTGGGAGTCCGAGGCAGGTGGATCACCTGAGGTCAGGTGTTTGAGACCAGCCTGGC
3773  CAACATGGCAAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGGGCACCTATAATCCCAGCTACTTGGGAGGCTGAG
3865  GCAAGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACTCCAGCCTGGGGGACAAGAGGAAAACTCC
3957  ATCTCAAAAAAAAAAAAGGGAATATTAATGAAGTAAAGTACATGTGATCTGCCATGGCCAGGGACAGGAATGCCATGGGGCCTGCAGCCGTCA
4049  CTAGCTGATGGCCCTTCTTTTTGCAGAATCAGATCCTGTCGCTTGGGGATCTCTGCCATCTGTGCTTTGGCTTCATGGTTCTCCTTGCCAGC
4141  AGCATCTTCTCTTCTAGATCTTTCCTACCCTTTAGAGACCACTTGAAATCCCATATTGTCTGAAGCTATTTAAGTCCACAGAAACTTTTCCC
4233  CCCACTGTCTCAATTCCTTTCCTACTGCCTGTCTGCACCGTGCACATAAACACTTGAGTATGTGGTCTTGGCTGTTCACGACCTACTTCTTA
4325  GGCTTCTTGCACGCAGGCATCCCGCCCCGTGCTGTGGTCCTGAGAAGGGCTGGCTTTGAGCCTCTGTTCTCCCACCCACCTGCCCACCTACA
4417  CATGCACAAAATCCCTTTCTTGCTAGGTGCTAGGGTTGAATACCCATTGCTTACCTTACTAATAGTAAAATTTTTACAAGCATTAGGTTATT
4509  TTCTTTGATTCATCAAGTAAATATTAATACTGTTTGGAACATGTGATAGTCCCAGCGACTAGATTTGTAAAAATATTTGCAGGATCAATGAT
```

FIG. 12B
1 of 10

```
4601  TTGGTTTGGCAGAAGTAGGTAATTTCTAAAATTAAAAAATGCAGGTAAAACAGGGACTGGAGAGGAGTATTTTTTCCTAGTGATTAATAAAC
4693  CTTTATTTTTCTTATTGTTTTGTTGTCTTACCCAGTTTATTTGGCGTAAATCTGAGAACTTACTTTTTCCATGAGCAAAGTTAGAGGTAAAC
4785  TTTAACAAGCAGTTAGACAGAGGTAATGACCTTTAGATTAAAAGGTTTTAGGTCAAGCTGTATAAGTTGACTTGTCGCTTAAGACATGATGA
4877  GCCTCTGTTTAACTGAAAGTCAAGCCCAGGACGCCTGCCTTTTCCATCAAAGACATGGGATTTGGGTGGCAGCTGACTATTGATTTCCAATG
4969  ACGATTCTTCTTCAAGTGGAGGTCTTTTTACCAGATGGTCCTGTTGGTGGGGACATTGTTAACCCTGCGATTAACCGACGGCATCTTCATCT
5061  GGCTTTTAAGCTCCTTGTATCCTGACTTGTTACACAGCTTACTTATGCTTGTGCGACTATGTAAAGTGACAGTATATGAGAAAGGTAGTGAG
5153  TAGTAAGAATGTTGGGAGACAATTTAAGCTACCATTCATATTTCATAAAAATTAGACTTTTGTGTCTGGTGTAAACAAACAGAGGACAGAGC
5245  TTGTATGAAAGGATAAAAGAGCGTTAAGGGTTACACGTCCATTAGGATAAAAAAACTAGAATATTTCTTTCTGAAACCTGAAGCCCAGGCCG
5337  GGCATGGTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGTTGAGATGGGAGATTGTTTGAGCCCAGGAGTTTGAGACCAGCCTGGGCAAC
5429  ATGGTGAAACCCCATCTCTATTTAAAGAATAAGGCTGGGTGTGGTGGCTCACACCTGTAATCCTAGTGCTTTGGGAGTGTGAGGCAGGTGGA
5521  TTGCTTGAGTTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCATCTGTACTAAAAATACAAAAATTAGGCGGGTGTGGTGGCG
5613  CCCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAAGCATGACAATCACTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCAC
5705  TGCACTCCAGCCTGGGTGACAGAGAGAGACTCCGTCTCAAAAAAATTAAAAAATTAGGCTGGGCGCAGTGGCTCACGCCTGTAATCCCAGCA
5797  CTTTGGGAGGCCGAGGTGGGCAGATCACGAGGTCAGGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAA
5889  AAAATTAGTTGGGCATGGTGGCAGGCGCCTGTAGTCCCAGCTGCTCGGGAGGCTGAGGCAGAAGAATGGCGTGAACCCGGGAGGCGGAGCTG
5981  GCAGTGAGCTGAGATTGCGCCACTGCACTCCAGACTGGGCGACAGAGCGAGACCTGTCTCAAAAAATAAAATAAAATAAAATAAATAAAAAA
6073  ATTAAAAAGAAAAAGAAAAGGAAACCTTAAGCCTAGTTATTGAGGTAGACAGGATGCTACCCCTGCCCTGTCATTTTATTTAAAAGAAGCAT
6165  TTAAGCCTAATGAACACGAGCAGTTCTAATGTCCGTTGGAGGGGAGGTAGCATTCACAGTTCATAGATTCATTTAGCAATTACTGATTGAGC
6257  ATCTTCTGTGTGTCTAGTTATCTATGCTCTTAGGCGCTGGGGATGTGGCAGTGAACAAGAACAGATGTAAATGACAAGAGATGGATGGTGGT
6349  GATGGTTGCACAATTGTGTGAATGTACTTAATGCCACTGAACTGTATACTTAAAAATGTTCAAAATGGCTGGGCATGGTGGCTCACGCCTGT
6441  AATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCACCTGAGATTAGGAGTTCGAGACCAGCCTGATCAACATGGAGAAACCCCGTCTCTA
6533  CTAAAAATACAAAATTAACCGAGCGTGGTGGCGCATGCCTGTAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATCGCTTGAACCCGGGT
6625  GGCGGAGGTTGCAGTGAGCCGAGATCCCGCCATTGCACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAAAAAAGTTTCAAATGG
6717  TAAATTTATGCATATTTTACCAGAATAAAAAAAGGCAGTTAAGACAAGTAAGATGCTGTGTCATGGGGCTAGATCAAGCACTTAGGGGTGGG
6809  GTGTTAGGGACTTTGAACGGGGCCTCTACCCTGCGGGAAGGGCTGAGCTGGAGGGATCTGTGGGCCCCTGATCAAGAAAGAAGCAGGAGCTG
6901  TAACCCAGCCTGGCTTTGGAACTTGAGGCTGCCCAGTGGAATCTGTTGTGTGTGACGGAAGGAGGCAGCTGCACTGGATGGGAGAAACTGGA
6993  GGGACTCTGTGGCTGCCAGGGCCAGCTGCAGGGCACACAGCTGCACTCTGAGGCTGGCACCTGCCTCCTTCACTTACCCAAGGCTTTTCCTC
7085  CTGTTTTTGTTTCCAGACGGCCATCCTGAGACGACAGGGTCGATACATATGCTCAACAGTGCCTGCGAAGGCGGAAGAGGAAGCACAGAGCT
7177  TGTTTGTTACAGAGGTAAGGCTCTTTCCTGCATTAATTTACATTTTGAAGTCATTTTCCCCTAACTGCCTCCTTTTCTTTAAATTTCAAAAT
7269  TGTCAAGGAAGTGTCAAAAGGGTAATTGTATTTCTATGATGGAAGTTCAAATAGAATAATGTGAATTTTTCAGACTCTGAAACTTGGACAGA
7361  AATGTCCACAGGGGCTATTTCTTTTTTACATTTTTATTATTTTTAAAACTTTATTTATTTGGAGGGGGCTATATCTGACTACAAAAAGTGAA
7453  TTCCACAGAATTTATCTCATGGACTTAAAATAAGCAGTAACTTGTAAATGAATTCACTGGAAATCTGTGGGAGGTCTTGTTATTGATACTGT
7545  TTTTAAGGGTGACACACACATTTATGTATCATTTATTTCATTTATACATTTATATTTGCAAGTTATTTTTGACTAGTATTCATGAAGTACTA
7637  GCTGATAATAAGCAGGGTCTATCGCTAGTCAATATATATTATTATATATATTGATTACTATATATATTCCTAATCAAGATACATTGATTAAT
7729  ATTATTTTTGTTTGAAAATGCAAATAAAATTATCTTATGGAAGAAAGATAAATTATTTACTTTTTTATTTTTATTTTTATTTTTTGAGACAG
7821  AGTCTTGCTCTGTTGTCTAAGCTAGAGTGCTGTGGAGCAATCTTGGCTCACTGCAACCTCTGCCTCTCCTGGGTTCAAGTGGTTCTCCTGCC
7913  TCAGCCTCCCAAGTAGCTGGGATTACAGGCGTGCACCACCACGCCTGGCTAACCTTTGTATTTTTAGTAGGGACAGGGTTTCAGCCTGTTAG
8005  TCAGGCTGGTCTCAGACTCCTGACCTCAAGTGATCTGCCCGCCTTGGCCTCCCAAAGTGCTAGGATTACTGGCATGAGCCACTGTGCCTGGC
8097  CAGAAAGAGAAATTATTACAATTTAGGTTGTTTGCTTTAGTTTTTCCCCTTGGAGTGTTTGTTTTTTCCTCCAGGTAATTTTAGGTAGGAAG
8189  GAATAATTAGATGTTTTAATTTTGTTTCTTTAAGTGCACCTTCATTTAAAAATAATGATTTTTTTTTAATCCTGGTTTTTCTAGTTGATATT
8281  TAGATCATAAATATGCTCATCAATAAAATTGCTTACTATAAGGAAGCTATAAATACTCTTATAAAGACCAATTAAATAACAATATTTAATTT
8373  CCATTGAGATTTTTGAAAAATTAAATATAAAATTAAAAAATTTTTAAGTGTGTTCCCTCATCTTTCTGAAGAAGTAACTTCCTGTCTTACCT
8465  CCTTTGCCACTATATTAGTAAACTTAATTCCAGACAAATACAGCCAGATATGTTTGTGAATGTAGTTATAAATGTCTTTTTAAGGCAGGTAG
8557  TGGCAAACTATGACCTGCAAGCAAAATCCAGCCTGTAGCCAATTTTTGTAAATAAAGTTTTATTGTAAAGCAGCCAAGCACATTTGTTTACA
8649  TATTGCCTATGGCTACTGTCACCATGCAACTCAAAGTTAAGTAGAGATAATATGACCCTCAAAGCTGAAAATATTTATGATCTTGCCTTTTA
8741  CACAAAAAGTTTGTTGACCTATGTTTTAAAGCATGTGGCAAAATTATTAATTGCTAACTCAGTTCTCCCAGTTGATTAAAAAAATATGGTTT
8833  TTTGAGGGAGAACTCTCCATTAAGTTATTTAATCACTGCAGGTTGAGCAATAGCTGCTTCATCCTATGCTGCTGGAGCCAACATAACTAAAC
8925  ACTTTTGGGACCCTTCCACTTCGGTGGAGTGAACATCACTTCCTCTTCATCCTCTGATCCAGGGAATGACCTAATGGCTTAAAACAAAGCAA
9017  AACAAAGCAGAAAAAAACTTCAAAAACCTTTCAGTGTAACTTCAAAATATTATTGAATTTCACCAGTTTGAAATGTTAATCGTATAATCAGT
9109  CAGTCACATTCCTGTCTTTTTGAAGGTACAGTTCTCAGGATCTGGCTTTCTGATGGAACTCCTTATCTCCTAGACTTTCTGCATCCCCAGAG
```

FIG. 12B
2 f10

```
9201    GGGTGGGGGTTGCTTGCCTAATTCTGTGCTTCTGCTTTGGAATTTAGCCAATGCCTGCTATGTAGATGGTCAACACTGGCTTGTTAAATCAA
9293    TGAATCTCTAAATACTTAGCCAGGTTCACTGTGGAGTTTTTTTTTGTTTTTTTTTTTTTTTGTATGTGCGATTCCTTATAATTATAAATATA
9385    TGAACATAATAAACTGTAATTATTTTCCCCATTTTGCCTAAGGTATTAAAATCTTGGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGTAC
9477    TTTGGGAGGCCGAGGCGGGTGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATTGTGAAACCCTGTCTCTACTAAAAATACAAA
9569    AAGTAGCCAGGCGTCGTGGTGTTCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGTAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGC
9661    AGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGGTAACAGAGCAAGATTCCGTCTTAAAAAAAAATAAATAAATAAATAAAAATCTGGG
9753    CTCCATTCATAAATGCATTTATAATATACAAATTTATGCTGGTTTATTTAGTTTTTAATTGGGCAGGTACATTTTATACTTCAGAATAGTTT
9845    TAATGTCCTGCAGGAGACTTGGTGGTACAATTTCAGTTCTAGAGCTTGTTGAAAAACCGCTGTAGCTCTTTGAAGAGAATAGGATCACACCC
9937    ACAAGACCACGTCGTACATAGTCAGGCCTGACTACTTAGCTGTGCCAGTGGACCTAGGGGCAGTAGTGGGAAAATCCAAAACTATGGATTAT
10029   TCTGAAATCTGACTGGTTTTGGAATGAAATTTGAGTGTGAATGTATGATACTCAAAATGTGAAAATGCTTTGAGAAATCTAAAAAAATCAT
10121   TTTCAAGGATAAAATAGCATTTTAAAATATTCTTAACTACAGAGTAAAAAAATCAAAAAGTTTGCCAGCTCCAAAATCTGACTTTGCGCTTA
10213   GCCCTTCTCTGCCATTTATTCCACCTCTCAGAATAATTTTAAAAACTTACATTACCTTCACATCACACATCACACCCTTGCCAGTAAAGTGC
10305   TTATTGTAGAGCCTGGCACAAAATAAGCACTTGCTGCTGTTGTAATTCTTACTTTCATGAGTTTGGAGGTGGAGGGGAGTTTGCTTCAGAAG
10397   ACCTTCAGGCATGGCCTGTGACACAGATAGAAAGGTTGACAGTAGAAACGATGAAGTGAAAGAAGCAGAATAGAGACGTTGTTCATGTTAGT
10489   TACAGCCAGAGCCAAATAAACTAGATAAGGCATAAATACATTTATCGACCTATGTTTTCATTGATTCACTCATGCTTTTTTTCCACCCACAC
10581   TGAAGTCTGCACATAAATAGTCAAGTGCTGACTGGACAGAAGATCATTTGAACTAAGCTGATTTTAACTAAAATTCTTTAGCTTTTGATTTT
10673   GACTCATTCATATTTGAAGATCTGGTGGATGTGGCATTATAATCTTAGGCTTTATGCACTATTGGAGGATCTTCATGGGTGGATGAACGACG
10765   GACACCAGGCTGGCCCATTGAATCTGTCATGTGGCGAGGTAGTTTGAATGCAGTCTGTGTAGGTGAGGTCACTAATAGAATCCTTGGCAGTG
10857   GAAGACTACATTAAACATTCTGTACACACACACAGATGAAAATGTCATGGTGCTAATATTAATTGGAATGCATATGTTTCTTCCTAATTATT
10949   TCACTCTTTTCCTTATGAATGGAAAGAAAAACTGAGGGCAGATTTATATGGTAATATTAACTAACAGAGCTAGCTAGTTCTAATATTAGTAA
11041   ATACCATGAAGGCGTCTTTAAGTTCACGCATCCTATTCACTTTTATGTCATAGTAGAGGCCACATGGCTTTTTAAGAACCAGGTGCTTACTC
11133   TGAAATAGTCCATTGTTAAGTAAATGTCACAAGGTTAGGTGAAGTTGTCTCCTTGTAAAACCTGCTCTCAGATCATTAATGATGATAACTTA
11225   AAGTGATACTACCCCCAAGGGTAATGTTTCAGTGGTTCAAAGTCTCAAGCTTCACAGGAATCTCTGGTGGTGGAATTTCAGGCATTAGCTCG
11317   TTGATGGGAAAAAATTTTTTTGATGTTTCTATGGGTATGCCTTCACAAACTTTCCCAAGTGTTTTCCAAAAATCACGTCTTTCTGTTTTTTA
11409   TTTTCTATTTTTAAAATCCCCCTTCAATAGTGCATCAAAACCTATAACTCTGACTGGCATCTTGTGAACTATAAATATGAAGATTACTCAGG
11501   AGAGTTTCGACAGCTTCCGAAGTGAGTAAGCTATATTATACACATAGGGAAAAGTCTTTGTACTTGAAATGCTTGGGGGGAGGTATGTAACT
11593   TCATATGCAATCAGAGTAATTGAGGAAAATATTTTTAGATGGTTTATGTGTATGTGGTGTAACTATTGTTTACAGGGCCTTTGATTGTAAGA
11685   CTCAAACATGCTACTTTGGTATTGATAGGTAGTAATAAACATGTGGATGGTTTAATTATGTCCAGTGGTTCTTTTCAGGGTACCACTGATAA
11777   ATAAATAGGTGAAATTTTTTCACTACAAAATATAAAACAAACAAAACGTAAGTGTACTGAAATATGCTGCAGTGCATTTTTTCCTTGAAAAG
11869   AATATTTTGAAGAAGATAATTATAAAGAGCATTTTACATTGAATAAATTTTATGTTTTTAAAAAAAGTAAATCAAGAAACAAGCATTTTCCC
11961   AGTTAAATTTTTTTTTTATCTCCTTTAATGTATTAACTTATTCTAGACTATACCAAAAGCAAGTGTATTTAGATTGAATAGTTGTGGCCAAA
12053   GTGAATCGCGGTAGCTAGGTATTGCCTTGACAGACTATCTTTTATAAAAGGTTCCATTGTGTGTTGCTTTAAAGGAATACCTGAGACTGGGT
12145   AATTTCTAAAGAAAAGAGATTTATTTATTTATTTTTATTTATTTAGTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATACAATGA
12237   CACGATCTCTGCTCACTGCAACCTCCACCTCCCAAGTTCAAGCAACTCTCCCGCCTCAGCCTCCCGAGTTGCTGGGATTACAGCCACTTGCC
12329   ACCACGCCCCGGCTAATTTTGTATTTTTAGTAGAGACAGGGTTTCACCGTGTTGGCCAGGCTGGTCTCGAACTCCTGAACTCAGGTGATCCAC
12421   CTGCCTCTGCCTCCCAAAGTGCTGGGATTACAAGCATTAGCCACCGAGACTAGCCAAGAAAAGAGATTTATTTCGCTTATAGTTCTGCAGGT
12513   TGTACAAGAAGCATAGCACTGGCTTCTGCTCAGCTTCTGTTGAAGACTTTTGTGCTGCTTCAAAACATGATGGAAAAGGTCAAAGGGAAAGT
12605   TGGCACTTGTGAAAAGAGACCAAAGAGGAGGAGGAAACTCACTTTATAACAACCCATTCTCTTGGGTACTAATCCATTTCAGCAACAAGTAA
12697   TCCCATCTTGCCAGGAACAAGAATTCACTCATTACTGTGAGAACAGCACCAAGCCTCTCATGATGGATCTGCCCCATAACCCACTAGGCCCA
12789   ATTATGCAACAACGGGGACCAAATTTCAATCTGAGTTTTGGTGGGAATAAAAACCATATCCAAACCATAGCAGTTGGCAAATTGAATTATAC
12881   TTGATTTTGGGAAAATTGAAAGCAAATAGTGATGGATTATGTTTTAAAACTAACCATCACCAATGAAAATATTACTTGAGACCTGATTAATG
12973   TATTTTTCTTTGCGTTTGTTACATCTTTGAGCTGGAACCTTTTATGCGGTTCTCAGTAGACCTAGCTGTTTGTTTTTCCTCCTTGTGTGGCT
13065   TTGCCACTCCTTAAGAATGTTCGGCCAATTCCCCGATTGCCTCTTTTTAAACCTCAGCCAGGAACACTCCCTCCTAGTATTATCTTCTCCAG
13157   ATGGGTAGCCCTTTAGTTCTATATTTACCCAATCCTCCCTTAGGGATTTTTTAATTCTTCCCATTGGATTGGCTTAACCCATCTTTGTTGAT
13249   CCTCTTGTTCATAGTCTCAGGGTTGAAAGATATCAGACTATGTCATGTCGTATACTTACTATCTAATAGACTGCTGGTACATTTTCTCTCTT
13341   GGCATTAATGAGAATTTCCAAATGTGTGATGAGAAAGAGAGGGAGAATTGTAACAGTGGTGAAGAACACAGATTCTGTGTTCTGCCTCAGAC
13433   GACTCTGCCGTCTGTCAGCCTGATTTTCCTCATCTGTTAAATGGCCTTAACAACAGTCACCATGATTAAAGGATTAAATGAGAGGGCACATG
13525   AGAAGTGCGCAGGGCCTAGCACGTCATACCCATTGAGTAAATGTAAGCTGCTTATTGGTATTGGTGTCTTGTTTTTGTTGTGGTTAATACCA
13617   TCATTGTTAATCGGTTTCAACGCAACAACTCAATCTTCCTTTTTTTCCTCATAAACTTTGTATTAAAAGTTATTCTACCAAGTCTTTGTTTA
13709   TTAAAAACTAATCCACTTTCTTATTTTTAGTATGCCTGCTAACTCCCCAGAAGCTATGCTGTCTTTTCCACATAGCTTTTTGGAGCTTTCTT
```

FIG. 12B
3 of 10

13801 ACTCAAGTCTCTTGGCTTACCCACCTTGAAAAGCAAGGGCATAGATGGTTTTATTCTTTGTCTGAATAAAGAAGCTGGGCCATCTTTGGATT
13893 TAGTAAAGCCCTGGCCCTATGATGGAGGAAGAAATGCAAAGCCTCTTCCTTGACTAGGCATTTCTAAAACAAATTTTCTTTGTTGTTTTTAT
13985 AGCAAAGTGGTCAAGTTGGATAAACTTCCAGTTCATGTCTATGAAGTTGACGAGGAGGTCGACAAAGATGAGGTGGGATACCTGCTTGCTGT
14077 TGCTTCTCTTTTCACTCTAGTTTAAACATCAATTTTACAGACTTAGAAGATTAGTTAGAAAATTACCGACATTTAGCCAAAACAGGCATTG
14169 GAGTGTTACATGAACGGGATAATTTTTTAAAAATGTTATTGATTGATTGGAATAAGGTCTCTGTTTCAACTTTACTGCTTAGCATTTCAT
14261 GTTCTCTTGGTTGTGTTTATTTGTTCTGAGATCATTTTCAAAGACTTGGATCAGATCTGGCTACATTGTTAAAAGATATCAAGATGACTTAG
14353 ACCTTGAATTTAGGTTGTTTTTCAACAGATCTCGAAACAGCTGCCAGCCAGTAGATTTAAATGGCTATTTCTTCAATGATTGCTTTTAGTGA
14445 AGTCTGATTTGATCAAGCCCACTCCCCCTATTCCTAGAGGAAAGCTCATGGCTAAAGAACTATATAAAGGGAGTAGGGCATTGAGATGAGTC
14537 TGCCCACTGAGTGAGGGAAACCTCACAAGAAGACAATGCCCATCTCTGCATTTCTCATCCTCCCCATTGATTGTTAAGTGTCCCATTGTGAG
14629 TTTAGGTTTTTCCTTCTTTAAAAAAATTGTCAGCTGAGCTATAACATTAGCCACTCATTAAGCAATGTGCATGTAGCAAATTATTTTTATTC
14721 CCCCCATCACTTTATCTCTCCTTTCTGCATTGCCTCAATTTCCTCCCTTGCTTTATTCACCTTTCCCTGAACTAAGCCTCTGGGAAGGTTTC
14813 CAGGAATGTGCATGTGCTTTTGTCCTCTGACTATAGGGGAGTGTCATTTGAAAACATTTTTTCGTGAAACCAGGCAAGACCTTCCAACGTGA
14905 GTGGTCAGTTGACCTACGTCCTTTTTGGTCCTTTTGTGGCTCATTAAACACTGACAAATAAAAATTTGGACAGGAGCTAGCTTTGCCTTTAA
14997 TGGAATAAAGTTTTCAGAAATGTAGGCGGGTCTCTCTCTTTCACCGCTAAGTGGACTTTTATGTGACTTGTAGGCATTGGTGTCAATGGGTG
15089 CTTCAGTAAATTTCAATGGACAACTTGGCACAAAGGGAATGACCTTCCCATTGACCAAACTCACAGCAAGCAACCCAGGTAATAACGGGAG
15181 GTTGACTGCTGGTGTTTTCCTTCTCTCCTAGGATGCTGCCTCCCTTGGTTCCCAGAAGGGTGGGATCACCAAGCATGGCTGGCTGTACAAAG
15273 GCAACATGAACAGTCCCTCAGCGTGACCATGAGGGTGAGGACGCACATCACTTTGCCCTCCCCTCTCACAAGCCCTTTCTGCCATAGAGCT
15365 CGAGAACAACCTCAAGATGAATGTCATGCTGTTCTTCCCCACAAAAGGGACATTGTCTGATTCCTAGGATGCTCCCCTGGTGATAGCACC
15457 CCCATTGGCACACTCATCCACCCACTTTCCCTCACTGTCTTCTGACCACCAGCATAAGGAGACCATCCCTGGGCTGGTGTGAAGGTGCAG
15549 ACACTGACATAGGCTTTCTTCTCTGTAATAACTGAAAAGTGCTCTTTGGTACCTCACAGAATGTCACCAAGGGGCTATCTGTCATGCCAATC
15641 CTGAGCACTTCTGTGCAAGTGTACTGCAGCAAAGTCAAGTAAAGCAAAAATTGAGGACGAGAAAAGAAATAGTTGCATAGAAGAGAAGGTT
15733 GCAGACAGAGAAGTCAAACCAATAGAAGAAGCTATTCAGGAGAAAAGTGGGACCAGAGGAACATCAGGATTAATAACAAAGGGAAGAGAAAC
15825 AAGGGAGTCAGTGAGATAAAAATTAAGAGGAAATGTGACTGTCATTACCCTAAGGCTGGAAAATCATTCAGCGTCATGAGGCAAAAAATAG
15917 TTCCCATTCTGTGAGCAAGAAACCCTGGGATTTTAGAGAAAGTTTCTGTCCTTCTGTGCTGCATCCCAAATTGGAAGTCCCTGCACTGCTT
16009 TTGGGTAGTTATGTAAAATCTCTGATTCCGTGGGTGAGAAAAATGACCCATGGATATTAGGGGAACCACCTCCTCAGAACTGAGATGCAGTG
16101 AGCTTCTTACATGCGATGGGAGTCTTGACCCCACAGTGACCTGGAGCATCAGCTAGAGTGAGAACGGAAACAGGTTTTATGTATGTATGTA
16193 GTCATAAGTGGGTTATTGATAGAGATTGTGACCCTCTTCATTTTGAAAGAATAGCTGTGTGTATATTTTTCTCTCAGTCATTTAAGAGACGA
16285 TTTTTCCACCTCATTCAACTTGGCGATGGATCCTATAATTTGAATTTTTATAAAGATGAAAAGATCTCCAAAGAACCAAAAGGATCAATATT
16377 TCTGGATTCCTCTATGGGTGTCGTTCAGGTAAATATGAAAAGAGTTTTACCATTATGTTTTCTTATCTGCAGTAGTGCTTATGTGTAAATTA
16469 GCAGATTTAAGCAAACACTTCCAAAAATGGCAATATGCATGGTAGAAATATAACATATAACTTTAAATGAGGCAAGCCTTGTTTTTCATCAT
16561 TGTAGAAGATGGAAGTGATAATGTAGAGGCAGAATTATGCTGTGGCAGGCAGGAGCACTCTGGCTCGGCCACTTTATAGCTGCGTGACCTTT
16653 AACAGGCTACTTAATTCAGATAATGAGAATGTTTCTTTAATACGGCAAATGAGTACATTGGATGAATCAGTGCAGGAAAATATTTAAAACAC
16745 TTCATAGTATCTCACTGGTGATTTTTATCGCTAGCATTGTAGTACCAGTGGCGGTGTAGATCAGTAAAGAGATTAGGTTTCAGCGCAGATTG
16837 AGTTCAAATCCCTGCTCCTCCACTTACCAACTGTGTAACCTTGGAGATGTTATTTAACCTCTCTGTACCTCAGTTTCTTCATTTGTTAAATA
16929 AGGATAATGGCAGTACCAAATATGGTTACTGAGAGGGTTCATTCATTACACATGTAAAAAGCTTAGAACAGTGCCAACAAATGGTAAGCATT
17021 TGGTCAGTATTAGATAGTTTTGTTATCATAGGGCTGTTGTACTTTTATATCATAGGGCTTATGTACTTATCCTTTAAAATTATTGTTAATTA
17113 AAGATAACACATGAATGTATTTTTCTTGTAAAAAATCAGCCAATACAGATAAAGTGAAAGTCCTTCTGGACTCCTCCCCTCCTTCAGTGTCT
17205 CTTTTCTGAGGGGAGCTACTACCAGTTTTGCATGCATCCTTCTGTAGCTTTTTAGCATTGTCTTTGGAAGAGAGTTGTCAATTTCCCTGTCC
17297 ATCATCTGTCCATCCATCCATCCATCCATCCATCTGTCCACCCCTCCATTCATCCAGCCTTGCCACTTTCAAGGAAGATTTAAGGCAGCAGC
17389 TTATAAGCATACACAGGACATGGGATAGCATAAATTTAAAGTGGGGGTGAAAGCAGAAAGATGAACAGGGGATTGGGATAAGGGTGAGAGAA
17481 AATAGAGTTAAGGAGAAAGCGTATGTTTTGAAGATCTAACACCTGCTGTGGGTGGGCCACCACCTGGGCTCTATGCTTTCTCACTTGGAGAC
17573 CTGTTTAGTCACGCAATTCACAGTGCACATGAGATAAAGGCATGATGCCTGTTTAGTCGACTCTAGAAGCACCCCTGACTTTAAAAAGAAGT
17665 TAAAGCAAAACTAAATGTATTTGGCAACCTCATTTTTTAAAGTAGGAAGTAATTATTTTTGTTTTATAAGAGAGTTTTGCTGCCTGTTTCTG
17757 GCCCAGGGACAGATGTTTATAAGTACAACTGCCCTGAGCTATCAATTAGTCTCCGGGGTGCATTTCAAAATCTAAGGTTCTGACTTCAATGG
17849 AAGTCTCTTCCTTCAAATTGTCTTTGCAGATGCAGCTGATGGTGTTTTCATTTAATAAAGTGTATCCAAGGCTTCAAAAAAGTAAAATAATT
17941 TGTTTTTATCTGTGTCTGTTTGTAAACTAAGCATCAAAAGTTGTGATTTAAGTGTTTTTAAAAATTATTACTTATGGATATTATAAAAAAAT
18033 TAGTTGACTGGTGCTGTGAATTAAAAAAAGTGCCTAAACTAAAAAATTTTGAAGCATTTTAGAACCCTTGAAATTTATTATACTTATTTTGC
18125 AGATGAGAAAACTGAGGCTCAGAAACAGAAATTTAGAATTGAGGCCTAATGTTTTTTTCTCCACTTTTAACTTTCTCTTTTCATGATTGTGA
18217 GTATGCAGGGAAAGGAGGAGAGAAAATTCATTTTGTTTCAAGCCTTTGACTTCTTCCCTGGTTCTTGCCTTGAAGTTTAAGTGGAATCCAAA
18309 GTGGCAATTACTGAGCCCACAGCAGACAGTCTGTGCACAAGAGTGTGTGGCTTTGCCAAGGGGAGCACTTGACTTTGCATTTCTAAGAACTG

FIG. 12B
4 of 10

18401 TGCTGCAGAATCACAGAGACTTTTGGGAGGGTTGCCCTGTCCCTGAGACCTCCACCAAGGAACTCTTAGAGAGAGTGTGGATAACCCAGTAG
18493 GATTTTAGTGGCTATGCGGGGGGCTGTCCTGCTGGCTCAGGTTAGTGGGAGTGTTTGATTTCATATCGCTCAGCCTGTCCTTACAGGGGATC
18585 TTGTGCCATGATCCTCAGAGCTGAACCTCTGTCTACTGCGGCCAACCTGGGGAGATTTTGCTCCCTGGAGGACATCTTGGAATGTCTGAAGA
18677 CTGGCATCTATTGGCTTAAGGCCATAAATTTCGCTAAACATTGTACAATGCATGGACCAGCCACTCACAACAAAGAATTGGCTGCCCAAGTG
18769 TCAGTAGTACCGAGATTGAGAAATCCTGGCCTAGTGCATGTTCATCTTCCGTCTGTTACTGCACATGGACTACTGTTCTTGTTCTGTGAGCC
18861 AGTCACCCTCTTGCAGGCATGAAAACTGGAGGCATGAGGCAAGGCCACGGACAGGGAGTCCAAATACCTTTTGGGATTCATAAAGGATGGGA
18953 AAGTTCCAGATAAGTAAGCCAAACATAGTAATAGATAATGGTTGGCTTTTAAAAATGTAATACCATACACTACTTCATTAAAAAAATAGGAG
19045 CTGAAGAAATATGAAAATTTTACATGAAATTTCATTTATTCAACAAATATTTTTCAAATACCCACTATGTGCAAGTCACTGTAGAGTCCATA
19137 GAGACTAAGGATGTGTAGCACTGACAAAAATGGGAGCACTGAGGAGGTTTCATTCCACTGCAGGGACACACAGTGAATCAGATGAGTATGTA
19229 AAGCAGGTAATGAGTCAGAAGGAAAAATAAAGCTTGCAGAAAGTGAAGCAGGGAAGGTGGACGGTAATGGGATTTCATGGGGGGGGGCTTTC
19321 ATGAGAAGGGGGCAAGCTATTTAAAATAGCTTGGTTCTAAATGCCAATGAGATATCACTCACCAACAAGAGAGAGTAATTATTTTAAAGCAG
19413 TTCTAATTCTTTTAAAGTATGTCTGTTTATCCTTTTTTCATTTCAGAACAACAAAGTCAGGCGTTTTGCTTTTGAGCTCAAGATGCAGGACA
19505 AAAGAGTTATCTCTTGGCAGCAGACAGTGAAGTGGAAATGGAAGAATGGATCACAATTCTAAATAAGATCCTCCAGCTCAACTTTGAAGCT
19597 GCAATGCAAGAAAAGCGAAATGGCGACTCTCACGAAGGTAGATAGGCTTGGCTTCCCCCAGGCACATACACACTCTGTGGGTGTCTTTATTT
19689 TTGCCAGGTGGGTATAAGAAGGAGACCTGTGTTACACAAGTACATGAGAGGTGGGACGGATAGGAGCTCTTTACAAATATCCTGTCAGCAAA
19781 GGTTTTGTCAGATTATAACTTACTTCCCTGACATTTCGTATATGGAAATCATGTAATGGGAAGAACCAAAGCTTTGGAGGCAGAAAGGGAGA
19873 CCTGCGTTTGAGTGCCATAAATACTGTATTTCAGCTGTGTAGCCCTGGGTAAACAACTTATGTTTTCTGAGCCTCAGTTGACTCACCTATAA
19965 AATGCGAATAAACATGAAAATTGCTGGGAAGATGGGAAGTGTAAATAAGAAAATGAATCTCAAGTATCTGGCATAGAATTTTACTGTATTAT
20057 AAAATATTAGTAATAATTAGAATGCATGGGAGCCTCAGATTAAATTGGTGAGAAAAATCTGGCTATGTTCTTGACAATTCATGTTTTACTTC
20149 AACCCTTAGGTGATTCCCAACCCTGGCTTCCCCTTAGAAGTACCTGGGAGCTTTTTAAAAATACCATTTACCTGGTCCCACAAAAGATTCTG
20241 ATTTAGTTGGTCTGGGGTGGAGCCTGGGCAGGTCTGACTTTTAGGGGGTCTCATGGACGTGTCCATGTGGGCTGTTGTTCATAGCTAGTGTC
20333 AGTTCTAATTGGACGGTGTCCATGCTATACCAGCTGCTCAGTGTTTTGACTTTCATCACTGAGCCTGTGGATCAGTATTTTTTCAAAGCACC
20425 CCAAGTGTTTCCCAGGAGCATCCAGAGTGGGGAACCACTGTGTTCATTTGAAGGCACCTAAGAGAAACGGCCTTCCTCCTCCTGTTTCAAAT
20517 GAAATGCTATGAATTACAAGTGATTCCGATAATCTGTTTTGCCATTTTAGATGATGAACAAAGCAAATTGGAAGGTTCTGGTTCCGGTTTAG
20609 ATAGCTACCTGCCGGAACTTGCCAAGGTAACATCGTCTTATATCTTCTGCTCTTCGTTGAATGCTGTTGAAGTATGTCTCATTTCACTGGTT
20701 TGTCCAGAATGGAATCTGTTGAAATCATAAAAATTACATTGTGATTACCTCTCTCTTTTTCTGACCTGATTACGAGGTGACGTGTACTCATG
20793 CAGTATGATTTCAGGTCTGTCTTCTAAAAAGTACCCTACAAAGCATTCTCCTTTTATTATTATTTTAAGTGTTTTTTTCCCTGATAATGCTT
20885 AACACTGCATCACAGGTACTGAAGAAATAACTGAAATATGCAGGCAGATGTTCTCATAATAGCATCGTACTTTCTATGTTGATACATGTGCT
20977 CTCCCTTACTCAGGGTAATAGACACGGTTCCAAAGAGGAAGGACCTGGTAATCTTGCCACGAAACCCGGGGGTTGCCTGAGTTACAGAAATT
21069 GTTTCGGGTCACTCTTACTGGAAAAAAAATAAGCTATTCCTGTGTCTTACAATTTGAGAAATTTAAAAGTTACTGAAAAGCACAAAGAAAA
21161 GCAAATCAACCATACTGCTACTTCCCAGATTAAATATCTATTATGATGTTGCCTTTTTAGCTTCCATATTCTTAAAAGATATAAAACATCGG
21253 TTATAGTTGAAGTTCTTTTTAAACACTGTCCTTATTCCCATTCTCTTCTTCTCCCACAACCCCAATCAGAAACAAGCACTATTAAAAGTTTT
21345 ACTTTCATTTTTATATCTTTACAAATAAATCTATCATAATGATATATACAGTATGATTTAGTATGTTTAAAAATGTTTTATAAATGCTAACA
21437 TACCATATGTATTCTGCACTTTAAAAAATTTTTAAATTTACCCCTTTTATTTGTACTATATAGACTTTTTATTTTAGCTGTTCTATTATTTT
21529 CATTTTTTTCTATTATAAACAAAGCTACAATGACTGTCCTTGTACTTGTGTCCTTGTGTGTATCTGAATGATTCTTTCTTAAATGAGAGAAA
21621 TATCTTTGTCTTCTCTAGCCCTTTGCACTCTTACTCTGTTACTGCCCTTCTATTCTTTTTTGATACTAGAGTGAAATGGCGACCCTCCACAC
21713 CCACATCTTAAACACTATAATAGAAACATGGTTTATCTATATAGGATTATAAATAGACCAGCATTCAGCATTGACCTTTATTTTAAGACAAC
21805 ATGGCTGTTCTCAAGTGTAAAATCTCCCTCCCTGGCTAGGGCTTTAGAGCATTGTTTTTCTTTAGGACTTGACTGCTACCACAGTATCTTTT
21897 TAGCACCTGCCTATTAAAGCTAATTTTAGTGCCACCATTGTAAACCACCTCCTAGTCTGGGAAGAGTTTTGGCTTGTGTGTTTGTGTTATGA
21989 ATGTCTGTGTATCATATTTTGCATTGAGATTTGCTTTTTTGTTCTGGATGTTTGGGGGTTCATAATTCTCAAAACAAAATATTTGTGCCC
22081 ATTTGGGTTTTAGTTTGTTGCAGCAGGTAATATATGTGATGCCATCTAGAATTCAGAAAGTAACCTTCTGCACTTACTGGGTGAACGGAATG
22173 GATACCTAGGAGAAGATTCATGTTATTTGAGCCTAATGTTGATTAATTAAAAATCTATGCTTTTTCCTATGAGGATATACAGGAACGGTCCC
22265 CTCCTTCTCTTACTACCCAGCCAATATAATTCAGTATTGTTTGATCCCAAGACCTAGGGAGATTTTTTAAGATATACATATATTTAATATAA
22357 ATGTATACATTTATGTATATATACTTTTTTATAAGTATAATTGTATATTTTGTCATTTAAAATATTTGGAACTATTTTTAAACTATGGGTTA
22449 CAAGTTAGGTTAAGCCATTTTAGTTGGTGAAATCAGTTTGATTTCAACCCCTGCTTCTTTTTGTTTTGTTTCTATTTAGTTTTTTATTCTTT
22541 TTATTGAGGTATAATTTGCAATAGCAGAATGCTCAAACATGAATTGTAGAGCTCACTGGAGTTCGCATTTGTACACTGATATAAGCAGCCCT
22633 CAGGCTAGCTTGTACCTGAGACCCTCTTTATTTTGACCTCCATCACCGTAGATTAGTTTTGACTTTTCTAGACCTTCCTGTGAATGGACTTA
22725 TACATGTACTCTTTGTGTCAGGCTTATTTAGCTAAACATGTGATTCACTTTAAGAAGTTTTTTTTAGGTCGGGCATAGTGGCTCATGCCTGT
22817 AATCCCAGCACTTTGGGAGGCTGAGGTGAGCGGATCTTTTGAGGTTAGGAGTTCAAGACCAGCCTTGCCAACATGGTATAAAACCCTGTCTC
22909 TACTAGAAATACAAAAATTAGCTAGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCATTTGAACCTG

FIG. 12B
5 of 10

```
23001  GAAGGCAGAGATTGCAGTGAGCTGAGATCATGCCACTGCACTCTAGCCTGGGTGACAGAGCAAGACTGTGTCTCAAAAAAAAAAAAAGGGTC
23093  CGTTTTAATGAAATAAAATGGAATGGAGAATATGAAAGTACACTGCCCTTAATAATGACATTATTTTTATATAAAATACTGTCATTATTAT
23185  TTTGGTGGCACCTGCCACCATGCCTAGCTAATTTTTGTATTTCTAGTAGAGACAGGGTTTTATACCATGTTGGCGAGGCTGGTCTTGAACTC
23277  CTAACCTCAAAAGATCCACCCACCTCAGCCTCCCAAAGTGCTGAGATTACAGGCATGAGCCACTACGCCCGACCTGAAAAAAAACTTTTTAA
23369  AGTGAATTACATAATTTTTTACATAAAATAATGTCATTATTAAGGGCAATGTACTATTTATACATATAGTGTGTATGTGTGTCTTGCATAGT
23461  GATATAAAAGATATTTGTTTTTCTTAGTGTGCTATTATGTATATTTATTTACTTTCATTGGTATATAATGTACCTATTTTGGGAGTTCATGT
23553  GATACTTTGATATCTGTATACAATGTGTGATGATCAAATCAGGATAATTGGGATAGCCATCACCTCAAACATTTATCTTTGTGTTGGGAATT
23645  TGAAACATTTCTTACCAGGAGTCATGGTCAAAACCTGAAAAATGAATCCTTGTTAGAGGCTTTTACTCTTTCCCCCTGGCTTTCAGGTGTTT
23737  TACAAATACTTTTATTTAGGAAGGTAGAAAGGTGGAAAGTAATTTTTTGAAGGGGAAAAGAATGAAGAAAATGGAGATGAGTTATTCACTCA
23829  GCACATGGGTATCTGTGGGCTTTGCCTTTTAAAGCCCAGCTTGGTGTCAGTGTGAGCAGCCCAGGCAGTAAGGGGAGACCTGTGTTCCCCAT
23921  CCCCAGCCTTGAGCAAAAATGCAGTTTTGGCTGTTTATCATCCCCCTTCAGGGTGTCTGAACTATTTGCACCGGTTGAGAAGGCAAAGAAGT
24013  TGACCTGATAACTGTTGGTCATCCCATTAGGAAGGATGGATTCCATGGTTACAGAATCAGAGACTGAAGTATGCAGAGGGAGGGGTGGGGAG
24105  AGAGAACTGTGCAAGGAGTTTACCCAGGGTATGAAGAGGTAAAGAGGTCAGTATCAGGGAAGGAAGGTGCAAGAAAGGGTCAGGCTGGGAGG
24197  CTGGGCCACAGTTCAGTAAGATTACAAAGAAGGGCCTAGAACAATGAGGGCAGGCAGAAGGTGGCTGAAGGTGTAATTTCATGGCAGGTTCC
24289  TTTTCTAATCAGCTCCTCTAACCTCCTTCATCCTGTTGCCCCGGCTTTGTTTTCCACTGTGACTAAGACATAGCCAAACAGGATATGACCG
24381  ACAGGAAGTTGTTTCAGTGCAAAAATAACTGATGTCTCATTCTGGAATATTATGGAAGGGCTCATTACTTACAGTGTGAGTGATGTAAACCC
24473  AGGTTTTCAGAATATTTTGTATAATCTTGGAGCTTATGTTTGTACATTTAGTACTGAACATCTGTATTGTTTTCTTATTAGAGAACACACTG
24565  TATTTACCCTAAAACTGGTTCTTTTCCTCCTATTGTCTATTATGGAACCAAACAATTTTTATTGTAAATGTAACAGTGTGTAGCATCAGTCT
24657  TATAAATATTTTAGTTTGATACACAAACCGTAGTTCAAGTTAGTTAATTGATTTCTTCCCTAGAAAGTCAAGGAGTAACATAATCAGGTTAT
24749  AAACTTCATTACTAGTTATTTAATAATTTATTTCTCTGGTTACATTTATATCTTAGGTGACATCAGAACATATATGTCACCTCCTTAAAGAT
24841  AGTGTGAAGAAAACCCACCTTATGTTTTCTTCCACAGCTTTTCTGTTTGTGAGCTTTTATTTTTGTACTCAAAGAATAGCATCCAACTTTTA
24933  CTTTGGTTTCCCCATGTGGTTCTGAAAGAGAAGTAGAATTTCTTCTAAATCCGGAATTGCTCACATCCTTTTACCTTTTAACTTTGTTTTAA
25025  GCAAATGAACTTATTGTTCCAGGTAAATCTTCCACAGTTGCATGCAGGGGAAAGTATGATGTCTCAGACTTTATAGTCTCATGGAGATGGAG
25117  TGAGGATCAAGGGCCATGCTCAGCAGAACTTGTCAGGACCCAGCAGTTTCACGGACACCTTTTCTTAATTTTTAAACCAAGTCTATAATAAG
25209  TGCTTTCTTCCCTAGATTCCAATCCAGAAAACAATATCATTGCACTATTATACAAAGGAGCTGGCTAGGCTTGTGTCTGTGGGGTCAGCTGG
25301  TGTTGCATTTCTGGGCCTCCTTTGTGAAGAGGATGAACTGATGGTCCTGAGAAGTTAGGTGTCTTGGAAGTAGTGGAAATAAATCATGATAA
25393  CTCTTTAAATTAAAGATTATATATTTTGGCCTCAAAACATTTTGCAAAGTCCTCCTATTCCAACCCAATCTGTTTAAATGACCCAACATTCA
25485  ACACATTGTTTCTGATAATTCATCCTCAGAATAAGATGCTGTTGGCCATAATCTTTGTCTCTAGATTGTTTTATCTACTCGCAAATAAATTT
25577  AAGACACAGAGTATGCCTAAAGCCTACAGCAGACTTTCTGGAAACTCTTGAATGTTTGGTCCATAACTACTTCTTAAGACAAAGAAGAAAAC
25669  CTTGTCAGGGTGTGTCATTAGTGCTTGAATGTAGGGTTTACAGGATGGGGTGGGGTGGGGGAATCGCCCTTGGTTTAGATGAATCATTCTTT
25761  TCCTTGTCTTCTCAGCAAACACCAGTTTCTACAGAGAACAGCTCTGCCATTGTGCATTTTCTGTCTCCATTTTCCTCTCATTCTCCTCTCCA
25853  CGAAACCCAGAGTAGTCAGTGGGCTTTGGGCAGGAAAGTGGCAACAGGGTGTCTGGGGAAAAGCCAGTTGGCTCTTCTTACCATCACAATAT
25945  AGACTGACCACAGGTTATTTTAAGAGCAGAGCTGGTTTCCATCACTCTGAGAAGTGCTCAACTACAGACTTTGGGATGATATTTGTTATAGC
26037  TGTATTTTCTCCACTCTTAGATTGTGAAAGTACATATTACAAGTATTTATTTTATTATCTTTACTAAAATTTTAATTAAAAAGAAGCGTGCT
26129  TGCCGCAATAAGTAAAAATACCCAAAGTTGTTTAAAGAAAAGTTCACCTTTTCCCTTCATCCTCCATTCCCACATTCCTGAGAACACTGAAG
26221  TTAATAACCGGTTGCAATTCCCCTTTCACCAAACTGATTGCTCATAGAAATATAGATAAACATATGTAAGGTTTTTAAGTTTTTTTAATAAA
26313  AATATGTTCATGATATATACATTATTCTGAAATTTTCTGTATCACTTAAAAATATTTCATAGATGTCCCTCTGGGCCAGTGGAAGATCTGGT
26405  TCCCCCTTACATACATATCAGCAAGCTGCATGATATTTCAACTATTGCTACTGCACAGTTTATTCAGCCATCTCCCTATTAATGAACATTTA
26497  AGGTTTTTTTTTCAGTTTTTAGCCTCTACAAACAGTACACAATAAACAACATGACATTAAATACTTGTGCTCTTATTTCAGTAGGAGAAATT
26589  CCCCAATGTGGAATTTTTAAGTCAAAGTTTATTGTGTTTTTAATGCTTTAAACATTGCCAGGTTACCGTCCCAAAAGGCTATAACAATTCAC
26681  ATTTCTGTTTCTCTGCATCTTCACCAGACGAGTGTAAAATGGTATTTTACTGTGCTTTCATTTATATTTTGCTGGTTATTAGTGATATTTTT
26773  CATATTTTCATATATTTATTTGCCATTTGTGTTTTTTCTTCCTGACTTGCTTGTTCACATTGTTTACCTTGTTTTCTTCTGTCTTGTGTAGT
26865  GTAATAGTTTAGACTCTGAAGCCAGGCAACCTGAGTTAGAAGCCAGGCCTCTATTTCATGATGTAGGTCTTTGGGCAAAGTACCTAACATTC
26957  ATGCCTTAGTGTTTTCTCTTTTAATGAGCAGGGATAATAATAGTACCTGCCTCCTAAGGTTGTATAAAATTAAAATGGGCACTTAGGGTAAT
27049  ATCTAGCAGGTAGATATTGGCTATTATCAATAGTAGCTCTTATCGTTACTATTCTTCCAGATACTGTTTCCTGACTCTGGGGCAAAGTCCTG
27141  CTACCCCTGAACCACATTTTTCTACCTCTTAGATTTTACTTGGTAATTCCATCAGCCACTGTTGGGCATCCTCTGTGTTTAATGCATCATCT
27233  TAGACCTTAGGAGGGATGGGAGGAACTTTAAGAAGCCGAATTTGCTTTTTATTTATCTTGTAGCAGAGCAATAGATGTATATTAGGTAGATT
27325  ACAAGCTTTTAGGTTATTTTTGCATCTAAAGCTGTCCCTTCTTTTCCAATAAATGATGTCTGTGGTAAAGAATATATCTGTTGGGTGTTAGT
27417  GACAAAATCAGAATGCTTTGTGTCTATTTTGGCTAGTAGTTAATTGTTTTTCTTTTATTGTGTCTGCATTCCTATTTGTTCTTTAATTATAC
27509  CGAGCTCATTAGCAGTTATTCTTGCTTTATTCATTTCTTATCTCCTAGCATAGTCAGCTCAAGACAACAAGCATCTTTCAGAAAGCCACTAG
```

FIG. 12B
6 of 10

27601 GAATTGCATCTACATTAAGAACCCACTCCTCTGCTCTAGCGTCTGAGAACATAACACAGTATTTGCCTTTGTTGAAGGGCTCAAGCGGACGA
27693 TTCAGAAGTAGAATAGATAACTGTTGGTGGTGGCTTGCAGCACCTAGTACTGATGGTTTTGTTAGGAAACACCACAGGCAGTATACAATGTA
27785 AAGCAAGTTCTTTGGGTTCAGTTAAATAGATTCCACTTGCACGTGTTCTCACTCTTTTGGTGTGAAAAATTAGGAAAGGTGATAGGCAGGAT
27877 AGAATAAAATGAAGTGGTCCCATCTCCCATATGGAGAGCGCCTGCCTCCACCACAGACACATGTTTTGCCCTGGAAGCATAAACAGAAGATT
27969 GCAGGAACGCCCCTTCACCTCCATAGCCTTCAGGCTCCCATCGATAGCATCAAGATAAACTTGGTGTGGCAACAGACTTGAGCCATCATCTT
28061 GTTTAACATTTTTACCTGGAAGTGAAAATGGAATCCAGAGAGGCTAAGTAGCTTGCACAGCTACTTAATTGAACTAGAACCAGAACCCAGTT
28153 TCTTAGTATTCTGGTGCCCATTTTATATAAACATAGACAGCTGATCATGGTAGTTTGGATCATTGCTAAAGACCTATTATATACAACAATCG
28245 TATGGCTAAGAAGATTAAGATGCTTATTTTCTCGTTATCCATGATTAGAAATGTAATAAGAGACTTTGTTCTGTCACCAACAGAAGAGACAA
28337 GACGCCTTTTCTGCTGTGGTTGTGAGTGCCCCATCAAGTGCTGAGATGTGACTCATGTCTTTGGGGAAAGTAGCAGATGAGAGAGTACATTC
28429 TATTCCGACTCCATAGACTTAACTGTGGAGACTGAGAGTAGTGGAGAGCTCAGACTGACAAGAAGAACAGAATATAGACTTAGAGGCACCAG
28521 GCATGAAAAATCATAAAGATGGAGATGACTTCTATCTTGTGAATGTCTGAAAAGCCTCTCTCTCTTTCTGTTCCCAAATCCTTCATCTCCAG
28613 CCCTCATCTCTGCCACCATCTGTCGGTTTTCTTCCGTGGCCTCTCTTCTGTCCACCTCTTCAATGTTGGGCCCCCAGGATTTTGTCTTCAGA
28705 CTCACTTTCACCCTCTCCCTCAGGACATCTCATCCACTCCCAAGGCTGGAACTCCAGCCTTCCTCTCAGCCTTGCGCTTAAGTTTCATCCCT
28797 TTAGCTGTCTATTGGATGTTTCCAGTTGGAAATTCCACAGGCATCTCTACCCAGGCGTCTCTACACAAAGATGATAAAAATATTATCATCTT
28889 TTTGGCAGATTTGTTTTTCCTGAGTTTTCTGTCGTGTTCATGGATTCACCATTCAGCAGGCTGCCCAAGCTAGAAATGTGGGATTTGTTAAT
28981 CGGCCCTGCTGTGAGACTGGGAGGCTGGTGTGATAACCCAAGAAAGACATTGTGGGCCTCTGGGTCTTATCCTTTTTAACCCATTGCCTCAG
29073 CCCTGCCCGAGTGATGCTTCTGAAATGTGGAACTCATTTTATTCTCTTACCTAAAAGGTTGCAGGATATTTAGATTCTGAATGAAATCCCC
29165 AATCCTTTTTTTTTTTTTTTAAGTGAAAGCAAGTTTATTAAGAAAGTAAAGAAATAAAGAATGGCTATGCCATTGGCAAAGCAGCCCTGTGG
29257 GCTGCTGGTTGCCCATTTTTATGTTTTTTTCTTGATGATATGCTAAACAAGGGGTGGGAAATCCCCAATCCTTGGCATTCAAATCCCAGACT
29349 CATTTCTTTTCACTTTTTTTTTTTTAATCATGCCCTGCCTTTCAGTTATGATGAGTGACTTGGTCATTGCTCAGGTGTGACTTGTCCCTTTGC
29441 ACCTGCTGTGCCTTCTCCTAGAATGCGCTTTTCTCCTTTCCTGGCCAAGTGTTCTTGTCCTTCCAAATGGGCCTTCCCTTGGGAGGTGGTTT
29533 CTGACGACCACCCCCTAGTCCAAGTCAGCTCCCACTGTACTTTAAACTTTCTCTTGTCTTCCTTATTACCTGTTGATATGCTCTCTCCCCAC
29625 CTGGTGTTCCTTGGGACTAGGGACTTCCTTCATTCACATTTCACATAACTTGAGGGCCTGGCTCATAAGAGGTGCTTAATGAATATTTATTG
29717 AATTAATTAGCATCTTGTCCTTCAAGATCAGCCATCATTTTCTCTATCTCATCATTCAAAATATATTCCTTCCTCTTCCCCTTCTTGCACCC
29809 AGTCACAGACTGGACTCTATTAAATCCTGTCTATCATCTGGGCTCATTTCCATCCTCAGTGTCTGTCTGTGCATCCTTTTCATTCAGCCAGG
29901 GATGTTCAGCTCGATTCTGCCCCTTCATTCCAAGCCTGTCCCATATTCCATTACTTTATGAAGCCTTTCTTGACACACAGATGCTTAATTAT
29993 TCTCTTTTGCTTTCTTTGTGTTGACTTTGACTCTGCCACTGGTTGTGAGCTTCAGAAGGGCAGGGATCTCACCTTCACTTCTTTTTCCTCCT
30085 AGTGCTTTCTTTGTGTGCTGCACACTCCCTGGCACACACAGCGGCTCTCCAACACGAGGCAGAGCTTTCCAGCAGCCTCAACCTTCAGGACT
30177 GGGCAGCTTTTAAATGTATTTGGGCACCTTTGCAAGAAAAGGATTGTGTTAAATGTAATTTTCATGTATCTTGTTATCAGAGTGCAAGAGAA
30269 GCAGAAATCAAACTGAAAAGTGAAAGCAGAGTCAAACTTTTTTATTTGGACCCAGATGCCCAGGTAAGAACTATCTAAATGTTTAATATTTA
30361 AAACCAAATGTGGGAGAGAAAATCATCGATGGGCTTATTTGTTTATTTGTTTGCTTTGTTTATTTTGGAAAAACAAGCAAAATAACTATAGA
30453 AAATTTGGGGAAAAGAGGAAAAATAAAAATGTATAATCTTATCACCATAGCATTACTATTGTGAATATTTGATATTCAATAGATGTTTGAAA
30545 ATTGGGAGAGATTTATTGAAAGACATTCTCAAGTTCACAAAGAACATCTAATTTACCTGTTAAAATAACCATCAGAAAACAACAGGTATCAC
30637 TGCAGTTGCCTGGGAGTCAGTGATAATTCCCGACTAGCCCAGGCTCAGGCTCAAATACAAACCTTTTCCATTTAACTCTAACGATAAGTACT
30729 TTTCTGTTTCCTCACAAACCTCATAACCATACGTATGTGTGTTTATATGTCTATATTTTTTATTTGCTTTTAAGAAGTTTTTGTTTATCATT
30821 GTAAAATATACATAATATAAAATTTACCATTTTAACCATTTTTAAGTGTACGGTTCAGTGGCATTAAATACATTCTCATTGTTGTACAACCA
30913 TTACCACCATCCATTTCCAGAACTTCTTCATTTTCCCACACGGAAACTTTGTATCAAATGATAACCTTCCTTTCCTTCTTCCCCCATCCCCT
31005 AGTAACCTCTGTTCTACTCTGTGAACCTGCCTATTTTAGGAACCTCATAAATGTGGAATCATACAGTATTTGTCCTTTGTTTCTGGCTTCTT
31097 AAACTTAACATGTTTTCAAGGTCAATCCATGTTGTAGCATGTGTCAGAATTTCCTTCCTTTCTGTGGCTGAATATTCCATTGTATGTATATA
31189 CTACATTTTATATATCCTTGTAAATCTGTTGATGGACACTTGGTTGGATACTTGATGGACATTGGTTTTGTTGTTCATGATCATAATTTTCA
31281 AGCTCTGTATTTTTTCAGTTCATCCATTGAGTAGGTATACCATCATGTCTTTTTTTTTTTGTCTTTTTTTTTTTTTTTTTTTGAGGCAGAG
31373 TCTTGCTCTGTCGCCCAGGCTGGAGTGCAGTGGTGCAATCTCGGCTCACTGCAAGCTCCGCCTCCTGGGTTCACACCATTCTCCTGCCTCAG
31465 CCTCAGCCTCCCGAGTAGCTGGGACCACAGGTGCCCACTACCACACCTGGCTAATTTTTTTGTATTTTTTGTAGAGACGGGGTCTCACTGGG
31557 TTAGCCAGGATGGTCTCGATCTCCTGACCTGGTGAGCCGCCAGCCTCGGCCTCCCAAAGTGCTGGAATTACAGGCGTGAGCCACCGTGCCCG
31649 GCCCATGTCTTTGACCATTGTTATAAACTATGTGTGTAACTACTATAAACCATAGAAACCGATTATATAATAGCAACACTATTGTGAGTAAA
31741 TAAGTGTATATAGCTTTTCCATATTTTATTCCGTTTCCTTTGGATGCATTTATGATGTTTTTTTAAAATAAGAGCATAACTTATTTATATGT
31833 TTACATTTTCTTTTAAAGAAGCTTGACTTCTCATCAGCTGAGCCAGAAGTGAAGTCATTTGAAGAGAAGTTTGGAAAAAGGATCCTTGTCAA
31925 GTGCAATGATTTATCTTTCAATTTGCAATGCTGTGTTGCCGAAAATGAAGAAGGACCCACTACAAATGTAATTTTTCATTTTAAAAATAAAC
32017 ATTAAAAAAAAAAATAGGCAGAGGTTTCAGATGTACCTTTACAGTGCAGCCTGGATAAGAAATCCTAGTCCCTGGTATCAAAGAGGTGCAGTG
32109 TTTGGATCAGGATATGGAGGTTGTTAGCCTGCAAGGACAGGATGTTCGTGATGGAAGATGAGGGTGGCAGGTTTGTGCTCAGCTTTCCAGGA

FIG. 12B
7 f10

32201 GACAGAGTTCATCTTAGATGCTTCAGGGAACGCAACTGTGTTTTCTATGGAACATACATTACGTAGCAAAACACATAAAGGTAAAATAATTG
32293 TTTTGTTTGTTTCCTAAAATGTTTTATAAGCTAATTCTCTGTATGCAGAAGGATGAGACTGTTTAGTAGTAATTTATGGCAACAGTCCTAAA
32385 TAGGTCTTGTCATTTCTCTTTTGATAGCAACATTCTTTTGTCCCTGTTTGAGCCATGGTGATCATTGGATTGTTTTGTTTTGTTCAGGTTGA
32477 ACCTTTCTTTGTTACTCTATCCCTGTTTGACATAAAATACAACCGGAAGATTTCTGCCGATTTCCACGTAGACCTGAACCATTTCTCAGTGA
32569 GGCAAATGCTCGCCACCACGTCCCCGGCGCTGATGAATGGCAGTGGGCAGAGCCCATCTGTCCTCAAGGGCATCCTTCATGAAGCCGCCATG
32661 CAGTATCCGAAGCAGGTGGGGAGTATGAGCCCAGCATTCCCACTACTCAGACTCACTTTGCATGCTACCTAAATGCACCAAAAATGCTCAAA
32753 TTAGACCTTGTAATGCACAAGTGGGGTCATTAGACTCTTAATTAATAGTATTTATTATTAGACAGTAAAAGCAAGCTGAGAAATAAGTGGCT
32845 TTTAATTTCCTTTCTTCTTTCTAAAAGTTCTCCTAGTTACCTCCCTCACCAAATAGACTTTTTGAGCAGATGATGAACTGTTTGTCAGCTAA
32937 CTAGTTTGGCACTGGGTGCTTTTTACTAGTTGTCCTGTTTCACTGTTCTTTGCTGTTTAATGTTCATGGGATTTGTTTAACGTAGCTGTGAA
33029 TTCTGCTTACTGAAGAAAATTGTTTGCCTCCTAGGGAATATTTTCAGTCACTTGTCCTCATCCAGATATATTTCTTGTGGCCAGAATTGAAA
33121 AAGTCCTTCAGGGGAGCATCACACATTGCGCTGAGCCATATATGAAAAGTTCAGACTCTTCTAAGGTATGAATGGCTTTTACGCTTTGGGGT
33213 GGTAAAAAGCAATCTGAAAAGAGGCCTTTATGTGATACTATAAATCCTTAATGAAATCAAACATAAGCCATATTTATACTCTAAAAGATGTA
33305 GAATATGCTACCTGTATTTACTCTGAACTTTATGTCTTGATTTTGAAGGAAATGTAGTCCTATGTGTAAATAAAATGATCCACTTGACTAGA
33397 CAGTTCTGACATCCTAAAATAATTTGCAAAGGAATTACCAGCTTAATAGTAAACTTTCTGTGTTAGAAGGTACATGTATGATATTCAAATAG
33489 AGTTTCTTCTATCTGTTAATTTGCCTCTTGGGTTCTGAAATTCTATTTTGGTCCACTTACACTTATATATGAGGCTGGAGACCAGGAGATGC
33581 CCTTGGCTCAGATGACCTGGCCAGCAGTGTCAGTGATTCAGGTCCACTTGGTTTTTGCTAGAAGGGGCAGAGTTTTAGGTGGAAGATGGAAA
33673 GAAAACATACATGATGTATGTTTGGTTTTTTTCAAAGTAGTGTTCATTACTTGGGAAGTGCCTAGGCATGGACATACCATAGAATTATTAAA
33765 TATTAGAGGTCATCTAGTCCAAGATGTGCTTTCATATTTCAGGACACTGAGGCTCACCAAGGCTTAGTGATCTGCTCAGTGTCTCATGGCTG
33857 GAGCTGCCAGGGGGCACTTGACTGCCACTTCGTAGCACCTTGTGCTACCTGGTTAGTGTAATCTGTTAAGTGCTATTATCCTTGCCAGTTTT
33949 ACATATTTTTAGTTATTTAAAAAAAAAATTGTGTGGTCATCTGAGAGGCCTGAATTCACTGACAAGGACTGGTCAGCAGTAGGAGTGGATAC
34041 CCCAGAGCCTGTGAGTGAAGTTTACTAAGTTGGATTCAGAGCTTCCTATCTTCACCTCTATGGGCGCCCATGCATCACAGCTGTGTCCACAG
34133 GATGCACGATGGCCATTGAGAAATGGATTTTGGAGTCAGAAGACCTGGGTGCTGCATGCTTAACTCATCTGGGTCCTTTGGACAAATCACAT
34225 CACCTCTCATGGCCTCCATATGTTCCTTCTGTGCATGAAGGATGATGTTACTTCTTGCCTCTGCCTTCCTCATAGGGACAGTGTTAGGATCA
34317 AACAGATCATGTATGAGTCAGTGCTGTGGGCACCATAAATCACAGAAAGCCCAGAAGACATCGTCATTTATTACAGCCCCAGTCAAGTAAAA
34409 GCCCATTTACCCAGGCACATTGGTTCCAACAGTAAGCCTTTTTGGCTGATGAAAGCTGTGTAAAGTTTGGTCTCTGGAGAGAAGCTGTTTTA
34501 TTTTTTTAAACCAAGTCTGTAAAACCTTGGATGAGAAGCTCTTTTAGCTCTTTTATGTTTTGATCAATAATCAATGAAGGCCCAATATAAGA
34593 TCTCCTCCCCCGACCGTGTATGCAACACATTTCCAAGGCCCATCCACAGCAACTTTGTTACTTCTGCCTGCCGCATGCATGGTTTGAAATTT
34685 GGCAGCTCATATTGGTGTAAAAATCACATATCACTGTAGGCTAAACTTACCTCTGCACACTCCTCCATGTCCACTGAGCATCTGCTGAAGTC
34777 TGCTTTTTCTTCATTTTTTTATGGAATGTAAAGCTCATCCATGTGTACATTATTCATGCATTTACTTTTCTGCCACCTCCAAAGCATTCAAT
34869 TAAAGCAGGAATTAAGGCTCAACTATCTTACTTTAGCACAGTTTTGGCAGAGATGTTACAGTGAGATGATTTTTTTCTGTCTGTCAAAGTTG
34961 TTTCTTCATGTTTTCCAAGATGGTCTAGAACATCATTTAGAGTAAATTTTCATTTTGGAGGAAATTTTTATGAAAAGTCTCTGTAGGTATCT
35053 CCTGTGAATAGAGGTTTTAAAAAGAAAAAGAAGGGGAAAAAAGCCCAAAGGGAAAAAATAAGTTTCTTACTCTGACTTTCACACATACTGTG
35145 TTCTATTTGCTCCCTTCATATGTCCGAGAGCTAAGTCCTCATTCACTGCAGAAAAGGCTTATTGATGTTTTATGTTTTAGCTTTAAATTTTA
35237 TGAAATTACTGCATTTTACTCCACAACATATTCATCATTGTTAGAACCAAAAAATCTTGAACCTGAAAATGTTTAAGTAAATTGACCCTGCA
35329 GCTAGGTAGGCCATTGTACCCTATAACTCATACACCTAAGACCCCACTAAGTCGCCCCACCCAGGGAGGCAAGAACATACCTCATTGGAGAA
35421 GGGGGGAGGTCCTAGATGATTCCCTGACATCTCTTCCAATTAGGATGTATTTTCCAGGTTACTCATGAAGGCAGCAGTTCCTACAGCATGAT
35513 TATGGTAGGATTATAAAGTTTTGCAACTGAAAGGGATCATGAAAAGTATAGCCTAGCCCCCATTTTCAGGTGTGGATAGTGGAGCTCAGAGA
35605 GGTTTGATGGCCTACAGAGTCATACAGCTGGATTCAGTGTGACTTAGAGACAGACAGCTCTGTCATCTAGATCCTTCTGGATCCTTCTAGAT
35697 CCTTCCTGTGGCTGATGGTACACACAGATCACCGGAGGTCTTGTCAGAACGCACGTTCTGATCCAGGAAGTTTGGGGCAAGGCCTGAGACTC
35789 TGCATTTCTACAGTGATGCTGATGCTATAGCACACTTTGGTTTTGGAGTACATTTCCCAAAATTGGTTTGACTTTGATATACTTATGTAAAA
35881 GACCCTTCAGTAAAAAAAAAAAAAAAAAATTAAGTAATGTAAAAGACCCTTCACTGTAACTCTGAGACAATCAAACTTTGCATTGATGACA
35973 AAGCACTATCACAGTACAGACAAATTTGGAAAAATTAAGCATTATCTTTTTTAGCAATAGGGGTATAGATGATTAATAATAGGTCTGCCTTA
36065 GGCTAAAAGCAAATAAGCTTCATTGTACACTGTGAATTAATCAATCAATCAACAACAAACCAAAATTGATCATTAGGAAGTAGAAATAGCAG
36157 TAATTCATGTTTAGAAAGCAAGAATATAGCATTGAGAACCCCAGCCTAACTGAAGTCAGTATCTTAATATCTTATCCCTTATATTGTGGGAA
36249 TCAGCATAAGCTAGTGCGCTTTAGGAGCTGTGAAAGCTTAGTATTTTAATTAGTGTTCTCATTTCAATCCTAATAATGTGATATATTTTGAT
36341 ATGGATACCAAATAGTAATTATTAATAACTCAGTAGACTTATAATAAGTAGCACTTAGTCATAAAGATGTATGAAAACCTCTGAAACAGCAT
36433 GTTTGTTGCCCGTAAAGACCAAAGAAGAGACATGGATTTTGGAAAGTCTTCGTTCTGTATTCTTTGAGAATCACTGCTAGAGAATGCGTTAA
36525 ATAAAGTACTCTGCACAGAGTGTGAATCCAGCCATATTCATTACTGTATGATGTGCTTTTCTATGCATCCAATGCTATGGTAAGGACTTATT
36617 TAGTGAGCATGTAAGGATGGCTCTTGAGGTCACAGTTCTTTCAGTGAGATGCAGTATCTCATAGCTTTATTCTCATGTCTTCAAGGTGGCCC
36709 AGAAGGTGCTGAAGAATGCCAAGCAGGCATGCCAAAGACTAGGACAGTATAGAATGCCATTTGCTTGGGCAGCAAGGTAAGGAACACCTTTT

FIG. 12B
8 of 10

```
36801  ATACCTTTTAAATCGATATAGATAGGTGCATGGATGGGTCAATAGGCCTATTCTGTTTGTTGTTCAGAGACAAAGAGGATTTGAATGTGTAA
36893  AACTGAGAAATACATAAGCCCAGATTTTGAAAAAATCATTTGGTAGAGTCACAGAGAGGATAGACACTGTCTGGAGAAGTGCTACCTGGAAC
36985  TGGCAGGGTGCACGGTAGTGTTAGCTGCAGAGCTGGGATTCAAGGACCCAACCACATGCCTCCAGCTGGAAGTCAGGGCAATCCAGTGAGGC
37077  CTGGGGTGATCTTTATCTCTTGACTCTACTTGTTAAGCATTTGACTTGTGTATATTGTTTCCTAAGCACAAGCCATTGGCTGGAACTGTTTT
37169  CTATGTAAATTGATTTAGTTGTCCTCATCCCCATAGATGTTTTCCATGTTTTTAGATAATGAGATTTCTGTTGGCTATAGCCAATGGAATAA
37261  TAATTAGACTTCTCATAGAAACTAGACTTAAATAATGAATTGATTTTGGTGTTTTGGAAACCCAGTTTAGAAATGTTGCTTTGCCATTTCAG
37353  GACATTGTTTAAGGATGCATCTGGAAATCTTGACAAAAATGCCAGATTTTCTGCCATCTACAGGCAAGACAGCAATAAGCTATCCAATGATG
37445  ACATGCTCAAGTTACTTGCAGACTTTCGGAAGTGAGTTTCAAGGTCTTATTTCCACACCTGAAAAATAGAAGCTGTGTAGTGGGGAGGGAGG
37537  AACAGGGGAGCAGTCACTTAGGTTGCTCGATTTAGACATCAGAGGGGATGGCAAATGAGCGTGAAGCATTTCCTCAAACCCTTGAGAAGAAA
37629  GATGGGGTGAAAATCAGAAGAATAACCAGTTAATTTGAATTCTGTAGAGGATGTTTTGGGTGGTGCTGTGAAGGGTGGACTGGGTAAGGATG
37721  AGCCTATGGTGGGGAGGAAACAGTTGAGGAACCTTGTCAAGAGGTGAGAAAGGACTCAGCAAAGCCACTGCAAGTGCAAACAGGAAGAAGGG
37813  GACAAATTCAAGTCGTGCCAAAGAGATACGATGACTGGGTCTTGGCTTGGGGTGGTAATAGTCTAAGATAAATAACTTGCAAGGTTTCTAAC
37905  TTGGAAGTTTCTGGCACCAGTTGTTGCTTAGCTTGTGGCAGCATTGCTCCACTCTGCCTCTGGCCTCACATGGTCCTCTTCCTGTGTGTCTC
37997  TGTCTAAATTCCCTTCTTAGAACACTAGTGATAAAGCATCAGGGCCAAGCCTAGTGACCTCATCTTAACTGATTGCATCTGCAGAGACCCTG
38089  TTTCCAAATAGGTCACATTTATAGGTACAAAGGGTTGGGACTTCACCATGCTTTTGGAGGACACAATTCAACCCATAACAATGAGGCAAAGA
38181  GGGAGCAAGGAATGTGTGCAACATCACAGGGCCGGCAGCTTCCCCAAGTCAGTCTCACCCGAGGGTCTGTGTTCTTAACCTCTATGCTGTTT
38273  TGCTGCTACATCCTAAAGAGTTCACTCTGAACCTTTGAAACTGATTTTCTTTGCTAGGGAGATGGGTCTTAGAATTTTTCTGGGGAAATTCT
38365  GGGAATGTGAAAGAGCTGAGGGCGCTAGAAGATGTGAAGTGAAAAGAATAGCTGAGAGCCAAATGCTAACTATTCTATGCCAAAGGTATCCT
38457  TGTTTTTTTTTTTTTTTGTGCATATCAAAATAGCAATCTTATCAGTTTGTCTAGAACTCAAGAATGATTGCTTAGCTTTCTTTAACCTTATT
38549  TTACCTTTTTCTTATCTGTCTTCAGTAGTAGGAATAGAAACGATATGAGTCATAGAAACAGGCTCAATAAGTTCTGAAAACACAGAGACGTG
38641  TTCCTAATCAGAATCCAATCACGTCCATGTCAGCAGGCGGCTTCAGCCTTCACAGCGACGTGAAATCCCTTGTCAAGAGGCTCAAAAAGGTA
38733  GAAAGGATTCTCAAGGTCTCTTTCAGTTATGTGATTATACAGTTTTTGACTGTCTTGATGTTTCCCCTGTTTGGAGCTTTAATGAGAAGTGC
38825  AACCTCAGTTTTGCTAACATGCAGCTAAGGTTGGCCTGTTCAGCAAAGCAGTGTGCATGCCCGCTGGGCTGATTTGGAATGAACCTTTCACA
38917  GCTCACGTAGGGAATTGGAGAAGGGGGAGAGGAGGATACTGGTGAAGGATGAGGCCTGCTGGGTTAGCCTTCCAGGGTTCCTGGACCATATA
39009  GGTGCCCCAAATTCCCAGTCACTATCTGACAGTTTTATGACCTGGTAAGGACACAGGTCTTGGCCAGGGAGTGCCCCTGGATCCCTATGAAT
39101  CTGTTATTCATGAAAGACTAAATAAAAGAATAGTACCCTATTTTACTTTTAAATCATAGAGGTTCTTTAGTTTACAAACATAATACATGTT
39193  CATTTTAGAAATTTTGAGAAATACAGAAGAATAAAAGGATGAAAAAAGGTTTACTACTAGTTTTAACCTTCGTGGTGAACTTTTGGAGAACT
39285  TTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCCCTCAGGCTGGCGTACAGTAGCACGATTTCAGCTCACTGCAACCTCCGCCTCCCGAG
39377  TTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTATAGGCGCCCACCACTACGCCTGGCTAATTTTTGTATTTTTAGTAGAGA
39469  TGGGGTTTCACCATATTGGCCAGGCTGATCTCAAACTGCTGACCTTGTGATCTGCCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCAT
39561  GAGCCACCGTGCCCAGCCTAGGGGGGAACATTTTTTTTACGTTTTATTCCTTTACATTTTATTTTAGTTTATCTTATGTAGCTATGATCAT
39653  ACTAAATATGTAATATTTCCCTGCACAACTCAAGTATTTTCTGAAAGTGTTATATATACATTTTTATAGACATCATTTTTAATGCATAAATA
39745  TTATAATAGTCCATTGAGATAGACCATAGATTATTTAACTCTTCCCCCATTTTTTGACTTTTTTTTTTTTTTCCGAGATGGAGTCTCGCTCT
39837  GTCGCCAGGCTGGAGTGCAGTGGCACCATCTCAGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGT
39929  AGCTGGGACTCCTGAGTAGCTGAGTAGCGCATGCTGCCACGCCCCGCTAATTTTTTTTTTTTTTTTTTTTTTTTTTGTATTTTAGTAGAGA
40021  TACTAAATATCTCACCATCTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAGGCAATCTGCCCGCCTCAGCTTCCCAAAGTGCTGGGATTAC
40113  AGGTGTGAGCCACCATGCCCAGCCATTTTTTGACTTTTAATGTGTTTCTGATTTTTCAGAATTATACCTATAAGCCACAGTTAGAATCTTTA
40205  AAAAAATCTTCTCTATTGGTAGTGGGTAATATATTATCATACATACTATATTATCATATAGTAATTATTGTCATTTTTTGAGTTTCAAGAAA
40297  ATTTCATTCTTACTAATTTTTTCAAAAACCAGTCACCTTTAGTTGGATAGATTTCAATATTTTCCTTCGCTCAACTACCATGCAACTCTTAA
40389  TAACCATGAGGTGGGTCTGCGTGTACTTAGGAAAGTGAATACACTATATTATTAAGGAAGAAAAAATATATCTGTATTACTATATTTTTTGA
40481  AAGAAAATATATATTTCTTTTGTATGTAAATGAAGAAATGGATAAGCAAGTAGCTATCTAGATGGAAAGATAGGCATAAAAATAGCTATTTA
40573  GGATATATGCCAAATAATCATGGTTATCTCTGAGGGATGGGTTGATGGGTGATATTTCACCTTCCACTTTATAACATTCTGTCATTTTTATA
40665  TGAACTTTTAAAAAACTAACACTTTTATATTCAGACAAAACAAACAATGAAGTTTTTTATATGTGATGGAGGTTGGAGCCCTGTCTCAGAAG
40757  TTACTTCCTAGGCTGGTTAGCTTGAGACTTCCCCACAGTGGCGGCTCCTCAGGGGCAGCCCCAGTGCATGGTCCTGTCTTCAGTGGAGGCTG
40849  GGGAGTGGGGCTTCACATGGTCACTAATTTGAAAGTGATGGGAGCAGAAAGCCTGTGGCCAGGCAGAAAGGAGCCCAGGGAAAACCAAGTGT
40941  GAGTTCTCTTCTGCACACCACTTCTTCATGCATGTGCTCAGCAGGAGGGCATTGGTGTGAAGGGTGTGCTCCAGGTGGCCAGTTAGAGACCC
41033  AGAAACCTGAAAACAGGGATCCGATGGTGACAGCATAGAAGACACAGCAGGATAAGTGAGGCCACGCTCCTCAATAAGTATTCAAAGAAACT
41125  TTGGTGCCCACTCCCCGTATTCTTCACAACAGAGTTAGGGGACGTGGAGGATTCCTTTTTCATTTTTTAAAAATCTTTGCATTGCTATTTTT
41217  CTTTCCTCTGTATATTTTACAGGAAATAATCTCATGTCAGTGGCCTGGGCACCGGCTTGGATCCAAAGCTATTGTTTCTACCCCCATGATTG
41309  TCTCAAAATGTTATTTAATAATGCATGAAAAAAATTTCTTCACGCTGTCTCAGTCTTAACAAAACAGCTGCCAAAGCTCATAAGCCACTTTC
```

FIG. 12B
9 of 10

41401 CTTTTTCCCTTGCAATAATTACCCAGGGATATGTTCCAAGATTTAGTAAGAAAGCGATTCTGTCCGATAGATGATATTGCTAACATTTTATA
41493 AGAAGAGAGACTTGGTACTTTGTATTTGATTTGTTCATGGTGGTATCTCATGGATAAGATGGTATCTCATCTTTTCCAACTTCTGCAGGAAA
41585 TGCGAAGACATGAAGGCAAAGTATAAAAATAGAACGTTTTCTTTAAAACGTAGACCTTTTTAATGGTACTACGTTGGATAGTTTAGGTAATA
41677 ATACTACTAAAGTTTTTGCGTATGCAGCTTAATGTGTCTGTGTTTATTTGTACACTCATCTTCTTTGCATCCAGGTTTTACAGTCTTACCCC
41769 GATTTCGCTCTGGTTACACTGCACTCAAGCCAAGTAGGGCTGCTTGACTTTCTCTAAACCCACTGGGGACTTCCCTCTGCCATGCTTTTCTC
41861 TCTGCCCAAATTGTGTCCCCTTCCTGCCTCATCAAGCAGCACATAAATCACAAACACATGCAGCATACACACTTCCCCTTTTCCTTTGTCTT
41953 TCTCAGGGAACTCTACTCATCTTTCAAAGCCCAGTCTGTGGCTCACTTCTGTGCTGGGAGTCCTGGAGGCGGTTACTTGGCTTCTCTGCCTG
42045 AGCCGCCTCCTCTTTTTAAGGGTGGATAATAACAGCCCCTGCCCCCTAAAACCGTGGTGGGGAATAAATGCAAAAGGCATTAAGGTGATTTC
42137 TCCCACCATGAATACTGATCTCATCCCGTGTTCCCTCTCGATAGATCTAGATACTCTGCCTTCTGGTAGAGGTTTGTACATACTCTGTGAAA
42229 GTGATTGCCCTCATATGCCGTAAGTAGCTTACAGTGTCTACTGGACTTTTGGCTTCTTGAGGAAAGAAATTATGTCTTGTTTGCATTCCTCC
42321 ATGGTCCTGAGTACATACATTGCAGCATATCCTAAGCACTTGATAAATGCTTATTGAATTTTCTTCTTAGACATAAACTCAGTGGTTTTTGT
42413 TGAAACAAAAATATCTCAAATTTCTTTCAATCATATATAGTTGTTTTTTTTTAAGTGACACCAAAGCTTTTAGGGAATATTTCCTTTCACAA
42505 AACACAGTTAGAAGATTAAACTCACCACCAATAGCAGTCCAAACATACCTGTATTGCCAGCTAATCATTTTAACGAGCCAATACAGGAAGTC
42597 AGGAAGGGAAGACCGGCTGCAGAAACACTTAGATAAGGACCCCAAATCTGTTGGCATGGGAGGACTGCTAGTTGATGATACCATTCCCATTT
42689 CCTCTGTGGGAATTGTTGAGTCAGCAGAAATGGATGGGCAGTGGGAAGGGAAAATTTTCCTAAGAGAGAGTTTGAGCCTCACTTCTACATTC
42781 ACACAGAGACAGGAGCAGTTCCCAGAGGCCAGGCATCCTGCAAGTGTCTGTATTGCATGCTTACTTAATTCGTGTAATTTTAAGATGAGTTT
42873 TCATGTTCAAGGATTATTTTATAAATTTTGCATAGAATATAGGTACTCTTTAGCAAAACAAAGCAAAAAAACCAAAACTATTCTCAGTCATG
42965 AAAGAATTCAGTTTGTGTAACACGCACACAACCACCACTTTGGAAGTGCATAAAAAGGCAGTAAAATCTTTATTGCCTGTGAGTGTTTGATG
43057 TCTAATAAACCAGATTCAACATAAACCATAAACTTTTGAAATGGGTTTGAGATTGGGTTTTTAAAAACTTAAAGCTGGCAAAAAAAAAAACA
43149 ACTTTTAAAAGCCCATGTGCTACATAATATGGAACTAAACTCAGAAATGTGCTTGGAAACACATGGAAAGAACGTCTTTACAGAAGCAGCAA
43241 CTAGAAGTAAAATCTCTCAGCAGAGGGAGGAAATAGAATAAGAAATAACTATAGTTAGGCACAGAAGGACACAATACACTATAGGAAGATTT
43333 CCAGTGAAGATCATTTAATTAAAATATGTTGCTTAGAAACGTATTTAATTGTGTTCCACCTCTCTCAAAAATTTATATGTGGAGGATGTTG
43425 GAGTGATCTTAAAAATGGTGATGAAGATGCCTGTTCATTCATAGGTGGAAATAATTAGGAGGGGGTGAAATCCATTACCCTTGCATACTTAC
43517 TTATATTTAAAAGTATAATTTGTAATAAA

FIG. 12B
10 of 10

```
hCLASP4     ----------------------MFPMEDISISVIGRQRRTVQ---------------- 20
hCLASP5     ----------------------MTHLNSLDVQLAQELG-------------------- 16
hCLASP3     --------------------MAERRAFAQKISRTVAAEVRKQISGQYSGSPQLLKNLNIVG 41
hCLASP2     ---------------------MLLFPYDDFQTAILRRQGRYICS--------------- 23
hCLASP7     -----------------MAASERRAFAHKINRTVAAEVRKQVSRERSGSPHSSRRCSSSL 43
hCLASP1     MSFRGKVFKREPSEFWKKRRTVRRVIQEEFHRFSSQEKPRLLEPLDYETVIEELEKTYRN 60
                                        ..:      .

hCLASP4     --------------------------STVPEDAEKRAQSLFVKECIKTYSTDWHVVNYK 53
hCLASP5     ------------------------------------------------------DFT 19
hCLASP3     N------------------ISHHTTVPLTEAVDPVDLEDYLITHPLAVDSGPLRDLIEFP 83
hCLASP2     --------------------------TVPAKAEEEAQSLFVTECIKTYNSDWHLVNYK 55
hCLASP7     G-----------------------VPLTEVVEPLDFEDVLLSRPPDAEPGPLRDLVEFP 79
hCLASP1     DPLQDLLFFPSDDFSAATVSWDIRTLYSTVPEDAEHKAENLLVKEACKFYSSQWHVVNYK 120
                                                                    ::

hCLASP4     YEDFSGDFRMLPCKSLRPEKIPNHVFEIDEDCEKDED--------SSSLCSQKGGVIKQG 105
hCLASP5     DDDLDVVFTPKECRTLQP-SLPEEGVELDPHVR---------------DCVQTYIREWLI 63
hCLASP3     PDDIEVVYSPRDGRTLVS-AVPEE-SEMDPHVR---------------DCIRSYTEDWAI 126
hCLASP2     YEDYSGEFRQLPNKVVKLDKLPVHVYEVDEEVDKDED--------AASLGSQKGGITKHG 107
hCLASP7     ADDLELLLQPRECRTTEP-GIPKD-EKLDAQVR---------------AAVEMYIEDWVI 122
hCLASP1     YEQYSGDIRQLPRAEYKPEKLPSHSFEIDHEDADKDEDTTSHSSSKGGGGAGGTGVFKSG 180
              :: .                :* .  ::* .

hCLASP4     WLHKANVNSTIT--VTMKVFKRRYFYLTQLPDGSYILNSYKDEKNSKESK-GCIYLDACI 162
hCLASP5     VNRKNQGSPEIC--GFKKTGSRKDFHKT-LPKQTFESETLECSEPAAQA--GPRHLNVLC 118
hCLASP3     VIRKYHKLGTGF--NPNTLDKQKERQKG-LPKQVFESDEAPDGNSYQDDQDDLKRRSMSI 183
hCLASP2     WLYKGNMNSAIS--VTMRSFKRRFFHLIQLGDGSYNLNFYKDEKISKEPK-GSIFLDSCM 164
hCLASP7     VHRRYQYLSAAY--SPVTTDTQRERQKG-LPRQVFEQDASGDERSGPEDSNDSRRGSGSP 179
hCLASP1     WLYKGNFNSTVNNTVTVRSFKKRYFQLTQLPDNSYIMNFYKDEKISKEPK-GCIFLDSCT 239
              : :               .::     *    : :    .   :   .   .

hCLASP4     DVVQCPKMRRHAFELKMLDKYSHYLAAETEQEMEEWLITLKKIIQINTDSLVQEKKETVE 222
hCLASP5     DVSGKGPVTACDFDLRSLQPDKRLENLLQQVSAEDFEKQNEEARRTN---------RQAE 169
hCLASP3     DDTPRGSWACSIFDLKNSLPDALLPNLLDRTPNEEIDRQNDDQRKSN---------RHKE 234
hCLASP2     GVVQNNKVRRFAFELKMQDKSSYLLAADSEVEMEEWITILNKILQLN-----FEAAMQEK 219
hCLASP7     EDTPRSSGASSIFDLRNLAADSLLPSLLERAAPEDVDRRNETLRRQH---------RPPA 230
hCLASP1     GVVQNNRLRKYAFELKMNDLTYFVLAAETESDMDEWIHTLNRILQISPEGPLQGRRSTEL 299
                      *:*:          .   ::    .   :

hCLASP4     TAQDDETSS----QGKAENIMASLERSMHPELMKYGRETEQLNKLSRGDGRQNLFSFDSE 278
hCLASP5     LFALYPSVD----EEDAVEIRPVPECPKEHLG-------N------RILVKLLTLKFEIE 212
hCLASP3     LFALHPSPD----EEEPIERLSVPDIPKEHFG-------------QRLLVKCLSLKFEIE 277
hCLASP2     RNGDSHEDD----EQSKLEGSGSGLDSYLPELAKSAREAEIK---LKSESRVKLFYLDPD 272
hCLASP7     LLTLYPAPD----EDEAVERCSRPEPPREHFG-------------QRILVKCLSLKFEIE 273
hCLASP1     TDLGLDSLDNSVTCECTPEETDSSENNLHADFAKYLTETEDTVKTTRNMERLNLFSLDPD 359
                      .       :                              :   :   : :: :

hCLASP4     VQRLDFS----GIEPDIKP-FEEKCNKRFLVNCHDLTFNILGQIGDNAKGPPTNVEPFFI 333
hCLASP5     IEPLFAS----IALYDVKERKKISENFHCDLNSDQFKGFLRAHTPSVAASSQARSAVFSV 268
hCLASP3     IEPIFAS----LALYDVKEKKKISENFYFDLNSEQMKGLLRPHVPPAAITTLARSAIFSI 333
hCLASP2     AQKLDFS----SAEPEVKS-FEEKFGKRILVKCNDLSFNLQCCVAENEEGPTTNVEPFFV 327
hCLASP7     IEPIFGI----LALYDVREKKKISENFYFDLNSDSMKGLLRAHGTHPAISTLARSAIFSV 329
hCLASP1     IDTLKLQKKDLLEPESVIKPFEEKAAKRIMIICKALNSNLQGCVTENENDPITNIEPFFV 419
              : :           .:    : .      : .. :.  :          . :.   * :
```

FIG. 13
1 of 6

```
hCLASP4         NLALFDVKNNCKISADFHVDLNPPSVREMLWGSSTQLASDGSP---KGSSPESYIHGIAE 390
hCLASP5         TYPSSDIYLVVKIEKVLQQGD----IGDCAEPYTVIKESDG------GKSKE-KIEKLKL 317
hCLASP3         TYPSQDVFLVIKLEKVLQQGD----IGECAEPYMIFKEADA------TKNKE-KLEKLKS 382
hCLASP2         TLSLFDIKYNRKISADFHVDLNHFSVRQMLATTSPALMNGS------GQSPSVLKGILHE 381
hCLASP7         TYPSPDIFLVIKLEKVLQQGD----ISECCEPYMVLKEVDT------AKNKE-KLEKLRL 378
hCLASP1         SVALYDLRDSRKISADFHVDLNHAAVRQMLLGASVALENGNIDTITPRQSEEPHIKGLPE 479
                . .   *:    *:.  :: .    : :          .        .. .    :

hCLASP4         SQLRYIQQGIFSVTNPHPEIFLVARIEKVLQGNITHCAEPYIKNSDPVKTAQKVHRTAKQ 450
hCLASP5         QAESFCQR-----LGKYRMPFAWAPISLSSFFNVSTLEREVTDVDSVVGRSPVGERRTLA 372
hCLASP3         QADQFCQR-----LGKYRMPFAWTAIHLMNIVSSAGSLERDSTEVEISTGERKGSWSERR 437
hCLASP2         AAMQYPKQGIFSVTCPHPDIFLVARIEKVLQGSITHCAEPYMKSSDSSKVAQKVLKNAKQ 441
hCLASP7         AAEQFCTR-----LGRYRMPFAWTAVHLANIVSSAGQLDRDSD----SEGERRPAWTDRR 429
hCLASP1         EWLKFPKQAVFSVSNPHSEIVLVAKIEKVLMGNIASGAEPYIKNPDSNKYAQKILKSNRQ 539
                     :  :        :   .  : :     . :

hCLASP4         VCSRLGQYRMPFAWAARPIFKDTQGSLDLDGRFSPLYKQDSSKLSSEDILKLLSEYKKPE 510
hCLASP5         QSRRLSERALSLEENGVGSNFKTS-----TLSVSSFFKQEGDRLSDEDLFKFLADYKRSS 427
hCLASP3         NSSIVGRRSLERTTSGDDACNLTSFR-PATLTVTNFFKQEGDRLSDEDLYKFLADMRRPS 496
hCLASP2         ACQRLGQYRMPFAWAARTLFKDASGNLDKNARFSAIYRQDSNKLSNDDMLKLLADFRKPE 501
hCLASP7         ---RRGPQ--DRASSGDDACSFSGFR-PATLTVTNFFKQEAERLSDEDLFKFLADMRRPS 483
hCLASP1         FCSKLGKYRRAFAWAVRSVFKDNQGNVDRDSRFSPLFRQESSKISTEDLVKLVSDYRRAD 599
                      .                        .: :::*:..::* :*: *:::: ::..

hCLASP4         --KTKLQIIPGQLNITVECVPVDLSNCITSSYVPLKPFE-KNCQNITVEVEEFVPEMTKY 567
hCLASP5         SLQRRVKSIPGLLRLEISTAPEIINCCLTPEMLPVKPFP-ENRTRPHKEILEFP--TREV 484
hCLASP3         SVLRRLRPITAQLKIDISPAPENPHYCLTPELLQVKLYP-DSRVRPTREILEFP--ARDV 553
hCLASP2         K-MAKLPVILGNLDITIDNVSSDFPNYVNSSYIPTKQFETCSKTPITFEVEEFVPCIPKH 560
hCLASP7         SLLRRLRPVTAQLKIDISPAPENPHFCLSPELLHIKPYP-DPRGRPTKEILEFP--AREV 540
hCLASP1         R-ISKMQTIPGSLDIAVDNVPLEHPNCVTSSFIPVKPFNMMAQTEPTVEVEEFVYDSTKY 658
                    ::  : . * : :. ..     :... :  * :          *: **     .

hCLASP4         CYPFTIYKNHLYVYPLQLKYDSQKTFAKARNIAVCVEFRDSDESDASALKCIYGKPAGSV 627
hCLASP5         YVPHTVYRNLLYVYPQRLNFVN--KLASARNITIKIQFMCG-EDASNAMPVIFGKSSGPE 541
hCLASP3         YVPNTTYRNLLYIYPQSLNFAN--RQGSARNITVKVQFMYG-EDPSNAMPVIFGKSSCSE 610
hCLASP2         TQPYTIYTNHLYVYPKYLKYDSQKSFAKARNIAICIEFKDSDEEDSQPLKCIYGRPGGPV 620
hCLASP7         YAPHTSYRNLLYVYPHSLNFSS--RQGSVRNLAVRVQYMTG-EDPSQALPVIFGKSSCSE 597
hCLASP1         CRPYRVYKNQIYIYPKHLKYDSQKCFNKARNITVCIEFKNSDEESAKPLKCIYGKPEGPL 718
                *   * * :*:**  *:: .     ..**::: ::: . *. :..:  *:*:.    .

hCLASP4         FTTNAYAVVSHHNQNPEFYDEIKIELPIHLHQKHHLLFTFYHVSCEINTKGTTKKQDTVE 687
hCLASP5         FLQEVYTAVTYHNKSPDFYEEVKIKLPAKLTVNHHLLFTFYHISCQQ------KQGASVE 595
hCLASP3         FSKEAYTAVVYHNRSPDFHEEIKVKLPATLTDHHHLLFTFYHVSCQQ------KQNTELE 664
hCLASP2         FTRSAFAAVLHHHQNPEFYDEIKIELPTQLHEKHHLLLTFFHVSCDNSSKGSTKKRDVVE 680
hCLASP7         FTREAFTPVVYHNKSPEFYEEFKLHLPACVTENHHLLFTFYHVSCQP------RPGTALE 651
hCLASP1         FTSAAYTAVLHHSQNPDFSDEVKIELPTQLHEKHHILFSFYHVTCDINAKANAKKKEALE 778
                *   .:: * :* :.*:* :*.*:.**  :  :**:*::*:*::*          :    :*

hCLASP4         TPVGFAWVPLLKDGRIITFEQQLPVSANLPPGYLNLNDAESRRQCNVDIKWVDGAKPLLK 747
hCLASP5         TLLGYSWLPILLNERLQTGSYCLPVALEKLPPNYSMHSAEKVPLQNPPIKWAEGHKGVFN 655
hCLASP3         TPVGYTWIPMLQNGRLKTGQFCLPVSLEKPPQAYSVLSPEVP---LPGMKWVDNHKGVFN 721
hCLASP2         TQVGYSWLPLLKDGRVVTSEQHIPVSANLPSGHLGYQELGMGRHYGPEIKWVDGGKPLLK 740
hCLASP7         TPVGFTWIPLLQHGRLRTGPFCLPVSVDQPPPSYSVLTPDVA---LPGMRWVDGHKGVFS 708
hCLASP1         TSVGYAWLPLMKHDQIASQEYNIPIATSLPPNYLSFQDSASGKHGGSDIKWVDGGKPLFK 838
                * :*::*:*::  .  :: :     :*::  .   .    .                ::*.:. * ::.
```

FIG. 13
2 of 6

```
hCLASP4     FKSHLESTIYTQDLHVHKFFHHCQLIQS----------GSKEVPGELIKYLKCLHAM 794
hCLASP5     IEVQAVSSVHTQDNHLEKFFTLCHSLESQVTFPIRVLDQKISEMALEHELKLSIICLNSS 715
hCLASP3     VEVVAVSSIHTQDPYLDKFFALVNALDEH-LFPVRIGDMRIMENNLENELKSSISALNSS 780
hCLASP2     ISTHLVSTVYTQDQHLHNFFQYCQKTES----------GAQALGNELVKYLKSLHAM 787
hCLASP7     VELTAVSSVHPQDPYLDKFFTLVHVLEEG-AFPFRLKDTVLSEGNVEQELRASLAALRLA 767
hCLASP1     VSTFVVSTVNTQDPHVNAFFQECQKREKD----------MSQSPTSNFIRSCKNLLNVE 887
                ..    *:: .** ::. **   :  :.                 :     *.

hCLASP4     EIQVMIQFLPVILMQLFR----------VLTNMTH------EDDVP 824
hCLASP5     RLEPLVLFLHLVLDKLFQLSVQPMVIAGQTANFSQFAFESVVAIANSLHNSKDLSKDQHG 775
hCLASP3     QLEPVVRFLHLLLDKLILLVIRPPVIAGQIVNLGQASFEAMASIINRLHKNLEGNHDQHG 840
hCLASP2     EGHVMIAFLPTILNQLFR----------VLT-RAT------QEEVA 816
hCLASP7     SPEPLVAFSHHVLDKLVRLVIRPPIISGQIVNLGRGAFEAMAHVVSLVHRSLEAAQDARG 827
hCLASP1     KIHAIMSFLPIILNQLFK----------VLVQNE------EDEIT 916
              . :: *   :* :*.                :          .:

hCLASP4     INCTMV-LLHIVSKCHEEGLDS----------YLRSFIKYS----------FRPEKP 860
hCLASP5     RNCLLASYVHYVFRLPEVQRDVPKSGAPTALLDPRSYHTYGRTSAAAVSSKLLQARVMSS 835
hCLASP3     RNSLLASYIHYVFRLPNTYPNSSSPG-PGGLGGSVHYATMARSAVRPASLNLNRSRSLSN 899
hCLASP2     VNVTRV-IIHVVAQCHEEGLES----------HLRSYVKYA----------YKAEPY 852
hCLASP7     HCPQLAAYVHYAFRLPGTEPSLPDGAPP----VTVQAATLARGSGRPASLYLARSKSISS 883
hCLASP1     TTVTRV-LPDIVAKCHEEQLDH----------SVQSYIKFV----------FKTRAC 952
              .   . . :     .                 .                 :

hCLASP4     SAPQAQLIH-------ETLATTMIAILKQS---------- 883
hCLASP5     SNPDLAGTHSAADEEVKNIMSSKIADRNCSRMSYYCSGSSDAPSSPA---------- 882
hCLASP3     SNPDISGTPTSPDDEVRSIIGSKGLDRSNSWVNTGGPKAAPWGSNPSPSAESTQAMDRSC 959
hCLASP2     VASEYKTVH-------EELTKSMTTILKPS---------- 875
hCLASP7     SNPDLAVAPGSVDDEVSRILASKLLHEELA-LQ---------- 915
hCLASP1     KE---RPVH-------EDLAKNVTGLLKSN---------- 972
                          :  .     .

hCLASP4     ----------ADFLSINKLLKYS----------WFFFEIIAKSM 907
hCLASP5     ----------APRPASKKHFHEELALQ----------MVVSTGMVKSM 910
hCLASP3     NRMSSHTETSSFLQTLTGRLPTKKLFHEELALQWVVCSGSVRESALQQAWFFFELMVKSM 1019
hCLASP2     ----------ADFLTSNKLLRYS----------WFFFDVLIKSM 899
hCLASP7     ----------WVVSSSAVREAILQHA----------WFFFQLMVKSM 942
hCLASP1     ----------DSPTVKHVLKHS----------WFFFAIILKSM 995
                          .  *                 ..   : ***

                  Cadherin Cleavage
hCLASP4     ATYLLEENKIKLFRGQRFPETYHHVLHSLLLAIIPHVTIRYAEIPDE---SRNVNYSLAS 964
hCLASP5     AQHVHNMDKRDSFRRTRFSDRFMDDITTIVNVVTSEIAALLVKPQKENEQAEKMNISLAF 970
hCLASP3     VHHLYFNDKLEAFRKSRFPERFMDDIAALVSTIASDIVSRFQKDTEM---VERLNTSLAF 1076
hCLASP2     AQHLIENSKVKLLRNQRFPASYHHAAETVVNMLMPHITQKFGDNPEA---SKNANHSLAV 956
hCLASP7     ALHLLLGQRLDTFRKLRFPGRFLDDITALVGSVGLEVITRVHKDVEL---AEHLNPSLAF 999
hCLASP1     AQHLIDTNKIQLFRPQRFPESYQNELDNLVMVLSDHVIWKYKDALEE---TRRATHSVAR 1052
              . ::   .: .  *  *. . : .    ::  :  .:     .  .     .. . *:*

hCLASP4     FLKRCLTLMDRGFIFNLINDYISGFSPKDP-----KVLAEYKFEFLQTICNHEHYIPLNL 1019
hCLASP5     FLYDLLSLMDRGFVFNLIRHYCSQLSAKLSNL---FTLISMRLEFLRILCSHEHYLNLNL 1027
hCLASP3     FLNDLLSVMDRGFVFSLIKSCYKQVSSKLYSLPNPSVLVSLRLDFLRIICSHEHYVTLNL 1136
hCLASP2     FIKRCFTFMDRGFVFKQINNYISCFAPGDP-----KTLFEYKFEFLRVVCNHEHYIPLNL 1011
hCLASP7     FLSDLLSLVDRGFVFSLVRAHYKQVATRLQSSPNPAALLTLRMEFTRILCSHEHYVTLNL 1059
hCLASP1     FLKRCFTFMDRGCVFKMVNNYISMFSSGDL-----KTLCQYKFDFLQEVCQHEHFIPLCL 1107
            *:    ::.:*** :*. :.    . .:.          .*    :::* : :*.***:: * *
```

FIG. 13
3 of 6

```
                                                      Cadherin EC motif
hCLASP4   PMAFAKPKLQR-----------------VQDS--NLEYSLSDEYCKHHFLVGILLRETSI 1060
hCLASP5   FEMNADTAPTSP--CPSISSQNSSSCSSFQDQKIASMFDLTSEYRQQHFLTGILFTELAA 1085
hCLASP3   PCSLLTPPASPSPSVSSATSQSSGFSTNVQDQKIANMFELSVPFRQQHYLAGIVLTELAV 1196
hCLASP2   PMPFGKGRIQR-----------------YQDL--QLDYSLTDEFCRNHFLVGILLREVGT 1052
hCLASP7   PCCPLSPPASPSPSVSSTTSQSSTFSSQAPDPKVTSMFELSGPFRQQHFLAGILLTELAL 1119
hCLASP1   PIRSANIPDPLTP------SES----TQELHASDMPEYSVTNEFCRKHFLIGILLREVGF 1157
                                    .        :.::  : ::*:* *::: * .

hCLASP4   ALQDN----YEIRYTAISVIKNLLIKHAFDTRYQHKNQQAKIAQLYLPFVGLLLENIQRL 1116
hCLASP5   ALDAEGEGISKVQRKAVSAIHSLLSSHDLDPRCVKPEVKVKIAALYLPLVGIILDALP-- 1143
hCLASP3   ILDPDAEGLFGLHKKVINMVHNLLSSHDSDPRYSDPQIKARVAMLYLPLIGIIMETVP-- 1254
hCLASP2   ALQEFR----EVRLIAISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVQRI 1108
hCLASP7   ALEPEAEGAFLLHKKAISAVHSLLCGHDTDPRYAEATVKARVAELYLPLLSIARDTLP-- 1177
hCLASP1   ALQEDQ----DVRHLALAVLKNLMAKHSFDDRYREPRKQAQIASLYMPLYGMLLDNMPRI 1213
           *:         ::   .:   ::.*:   *   *  *       :.::* **:*: .:  : :

hCLASP4   AGRDTLYSCA-----------AMPN-S----ASRDEFPCGFTSPANRGSLSTDKDTAYGS 1160
hCLASP5   ---------------------QL---------------CDFTVADTRRYRTSGSD----- 1162
hCLASP3   ---------------------QLY--------------DFTETHNQRGRPICIATDD-- 1276
hCLASP2   NVRDVSPFPVNAGMTVKDESLALPA-VNPLVTPQKGSTLDNSLHKDLLGAISGIASPYTT 1167
hCLASP7   ---------------------RLH--------------DFAEGPGQRSRLASMLDSDTE 1201
hCLASP1   YLKDLYPFTVNTSNQGSRDDLSTNGGFQSQTAIKHANSVDTSFSKDVLNSIAAFSSIAIS 1273
                                            . :

hCLASP4   FQ-NGHGIKREDSRGSLIPEGATGFPDQGNTGEN------TRQSSTRSSVSQYNRLDQYE 1213
hCLASP5   --------EEQEGAGAINQNVALAIAGNNFNLKT------SGIVLSSLPYKQYNMLNADT 1208
hCLASP3   --------YESESGSMISQTVAMAIAGTSVPQLTR----PGSFLLTSTSGRQHTTFSAES 1324
hCLASP2   STPNINSVRNADSRGSLISTDSGNSLPERNSEKSNSLDKHQQSSTLGNSVVRCDKLDQSE 1227
hCLASP7   --------GEGDIAGTINPSVAMAIAGGPLAPGSR----ASISQGPPTASRAGCALSAES 1249
hCLASP1   ------TVNHADSRASLASLDSNPSTNEKSSEKTDNCEKIPRPLALIGSTLRFDRLDQAE 1327
                 . :  . :   :           .            .       :.

hCLASP4   IRSLLMCYLYIVKMISEDTLLTYWNKVSPQELINILILLEVCLFHFRYMGKRNIARVHDA 1273
hCLASP5   TRNLMICFLWIMKNADQSLIRKWIADLPSTQLNRILDLLFICVLCFEYKGKQSSDKVSTQ 1268
hCLASP3   SRSLLICLLWVLKNADETVLQKWFTDLSVLQLNRLLDLLYLCVSCFEYKGKKVFERMNSL 1384
hCLASP2   IKSLLMCFLYILKSMSDDALFTYWNKASTSELMDFFTISEVCLHQFQYMGKRYIARNQEG 1287
hCLASP7   SRTLLACVLWVLKNTEPALLQRWATDLTLPQLGRLLDLLYLCLAAFEYKGKKAFERINSL 1309
hCLASP1   TRSLLMCFLHIMKTISYETLIAYWQRAPSPEVSDFFSILDVCLQNFRYLGKRNIIRKIAA 1387
           :.*: * * ::*  .  :  :   .  ::  :: :  :*:  *.* **:   :

hCLASP4   WLSKHFGIDR----------------KSQTMPALRNRSGVMQARLQHLSSLESS------ 1311
hCLASP5   VLQKSRDVKAR---------------LEEALLRGEGARGEMMRRRAPGNDRFPGLNEN--- 1311
hCLASP3   TFKKSKDMRAK---------------LEEAILGSIGARQEMVRRSRGQLERSPSGSAFGSQ 1430
hCLASP2   LGPIVHDRKS------------------QTLPVSRNRTGMMHARLQQLGSLDNS------ 1323
hCLASP7   TFKKSLDMKAR---------------LEEAILGTIGARQEMVRRSRERSPFGNPEN----- 1350
hCLASP1   AFKFVQSTQNNGTLKGSNPSCQTSGLLAQWMHSTSRHEGHKQHRSQTLPIIRGKN----- 1442
             .                        :        :       :

hCLASP4   ------------FTLNHSSTTTEADIFHQALLEGNTATEVSLTVLDTISFFTQCFKTQLL 1359
hCLASP5   --LRWKKEQTHWRQANEKLDKTKAELDQEALISGNLATEAHLIIILDMQENIIQASS-ALD 1368
hCLASP3   ENLRWRKDMTHWRQNTEKLDKSRAEIEHEALIDGNLATEANLIIILDTLEIVVQTVS-VTE 1489
hCLASP2   ------------LTFNHSYGHSDADVLHQSLLEANIATEVCLTALDTLSLFTLAFKNQLL 1371
hCLASP7   --VRWRKSVTHWKQTSDRVDKTKDEMEHEALVEGNLATEASLVVLDTLEIIVQTVM-LSE 1407
hCLASP1   --ALSNPKLLQMLDNTMTSNSNEIDIVHHVDTEANIATEGCLTILDLVSLFTQTHQRQLQ 1500
                    .     .     :: :.     ..* ***   *   **    .  .
```

FIG. 13
4 f6

```
hCLASP4         NNDGHNPLMKKVFDIHLAFLKNGQSEVSLKHVFASLRAFISKFPSAFFKGRVNMCAAFCY 1419
hCLASP5         CKDS---LLGGVLRVLVNSLNCDQSTTYLTHCFATLRALIAKFGDLLFEEEVEQCFDLCH 1425
hCLASP3         SKES---ILGGVLKVLLHSMACNQSAVYLQHCFATQRALVSKFPELLFEEETEQCADLCL 1546
hCLASP2         ADHGHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRSLIYKFPSTFYEGRADMCAALCY 1431
hCLASP7         ARES---VLGAVLKVVLYSLGSAQSALFLQHGLATQRALVSKFPELLFEEDTELCADLCL 1464
hCLASP1         QCDCQNSLMKRGFDTYMLFFQVNQSATALKHVFASLRLFVCKFPSAFFQGPADLCGSFCY 1560
                   .    ::   :   :  :   **   * : ::: * :: ** . :::  .: *  :*

hCLASP4         EVLKCCTSKISSTRNEASALLYLLMRNNFEYTKRKTFLRTHLQIIIAVSQLIADVALSGG 1479
hCLASP5         QVLHHCSSSMDVTRSQACATLYLLMR--FSFGATSNFARVKMQVTMSLASLVGRAPDFNE 1483
hCLASP3         RLLRHCSSSIGTIRSHPSASLYLLMR--QNFEIGNNFARVKMQVPMSLSSLVGTSQNFNE 1604
hCLASP2         EILKCCNSKLSSIRTEASQLLYFLMRNNFDYTGKKSFVRTHLQVIISVSQLIADVVGIGE 1491
hCLASP7         RLLRHCCSRISTIRTHASASLYLLMR--QNFEIGHNFARVKMQVTMSLSSLVGTTQNFSE 1522
hCLASP1         EVLKCCNHRSASTQTEASALLYLFMRKNFEFNKQKSIVRSHLQLIKAVSQLIADAG-IGG 1619
                .:*: *       :.... **:**   .:    .: * ::*:  :::.*:.      .

hCLASP4         SRFCESLFIINNFANSDRPMKATAFPAEVKDLTKRIRTVLMATAQMKEHEKDPEMLIDLQ 1539
hCLASP5         EHLRRSLKTILAYSEEDTAMQMTPFPTQVEELLCNLNSILYDTVKMREFQEDPEMLMDLM 1543
hCLASP3         EHLRRSLKTILTYAEEDLELRETTFPDQVQDLVFNLHMILSDTVKMKEHQEDPEMLIDLM 1664
hCLASP2         TRFCQSLSIINNCANSDRLIKHTSFSSDVKDLTKRIRTVLMATAQMKEHENDPEMLVDLQ 1551
hCLASP7         FHLRRSLKTILTYAEEDMGLRDSTFAEQVQDLMFNLHMILTDTVKMKEHQEDPEMLIDLM 1582
hCLASP1         SRFQHSLAITNNFANGDKQMKNSNFPAEVKDLTKRIRTVLMATAQMKEHEKDPEMLVDLQ 1679
                  ::.**      :: *  :: : *. :*::*  .:. :*  *.:*:*.::***** :**

                                                transmembrane
hCLASP4         YSIAKSYASTPELRKTWLDSMAKIHVKNGDFSEAAMCYVHVAALVAEFLHRKK------- 1592
hCLASP5         YRIAKSYQASPDLRLTWLQNMAEKHTKKKQYTEAAMCLVHAAALVAEYLSMLEDH----- 1598
hCLASP3         YRIAKGYQTSPE-RLTWLQNMAGKHSERSNHAEAAQCLVHSAALVAEYLSMLEDR----- 1718
hCLASP2         YSLAKSYASTPELRKTWLDSMARIHVKNGDLSEAAMCYVHVTALVAEYITRKG------- 1604
hCLASP7         YRIARGYQGSPDLRLTWLQNMAGKHAELGNHAEAAQCMVHAAALVAEYIALLEDQ----- 1637
hCLASP1         YSLKNSYASTPELRRTWLESMAKIHARNGDLSEAAMCYIHIAALIAEYLKRKGYWKVEKI 1739
                * :*..*  :*: * ***:.**  * .   :*** * :* :**:**:*

hCLASP4         -------------------------LFPNGCSAFKKITPNIDEEGAMKEDAGMMD---- 1622
hCLASP5         -------------------------SYLPVGSVSFQNISSNVLEESVVSEDTLSPDEDGV 1633
hCLASP3         -------------------------KYLPVGCVTFQNISSNVLEESAVSDDVVSPDEEGI 1753
hCLASP2         -------------------------VFRQGCTAFRVITPNIDEEASMMEDVGMQD---- 1634
hCLASP7         -------------------------RHLPVGCVSFQNISSNVLEESAISDDILSPDEEGF 1672
hCLASP1         CTASLLSEDTHPCDSNSLLTTPSGGSMFSMGWPAFLSITPNIKEEGAAKEDSGMHD---- 1795
                                        :  *  :*  *:.*: **.   :*     *
                                                                   ITAM

hCLASP4         ---VHYSEEVLLELLEQCVDGLWKAERYEIISEISKLIVPIYEKRREFEKLTQVYRTLHG 1679
hCLASP5         CAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPILEAHREFRKLTLTHSKLQR 1693
hCLASP3         CSGKYFTESGLVGLLEQAAASFSMAGMYEAVNEVYKVLIPIHEANRDAKKLSTIHGKLQE 1813
hCLASP2         ---VHFNEDVLMELLEQCADGLWKAERYELIADIYKLIIPIYEKRR---------------- 1677
hCLASP7         CSGKHFTELGLVGLLEQAAGYFTMGGLYEAVNEVYKNLIPILEAHRDYKKLAAVHGKLQE 1732
hCLASP1         ---TPYNENILVEQLYMCGEFLWKSERYELIADVNKPIIAVFEKQRDFKKLSDLYYDIHR 1852
                     :.*  *:  *  .  :  .  ** : :: * ::.: * .*
                ITAM                DOCK motif       DOCK motif                ITAM
hCLASP4         AYTKILEVMHTKKRLLGTFFRVAEYGQSFFEEEDGKEYIYKEEKLTGLSEISLRLVKDYG 1739
hCLASP5         AFDSIVNKDH--KRMFGTYFRVGEFG-SKFGDLDEQEEVYKERAITKLPEISHRLEAEYG 1750
hCLASP3         AFSKIVHQSTGWERMFGTYFRVGEYG-TKFGDLDEQEEVYKERAITKLAEISHRLEGEYG 1872
hCLASP2         ----------------------------DFFEDEDGKEYIYKEEKLTPLSEISQRLLKDYS 1710
hCLASP7         AFTKIMHQSSGWERVFGTYFRVGEYG-AHFGDLDEQEEVYKEESITKLAEISHRLEEEYT 1791
hCLASP1         SYLKVAEVVNSEKRLFGRYYRVAEYGQGFFEEEEGKEYIYKEEKLTGLSEISQRLLKDYA 1912
                                          * : :  :*:::**** :* *.*** **   :*
```

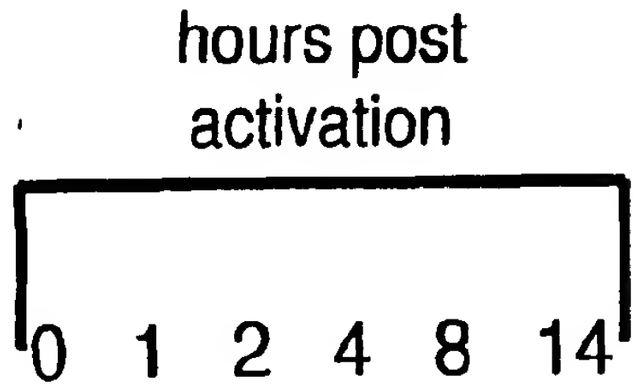
FIG. 13
5 of 6

```
                ITAM                                    ITAM
hCLASP4          EKFGTENVKIIQDSDKVNAKELDPKYAHIQVTYVKEYFDDKELTERKTEFERNHNISRFV 1799
hCLASP5          QCFGAEFVEVIKDSTPVDKTKLDPNKAYIQITFVEEYFDEYEMKDRVTYFEKNFNLRRFM 1810
hCLASP3          ERFGEDVVEVIKDSNPVDKCKLDPNKAYIQITYVEEYFDTYEMKDRITYFDKNYNLRRFM 1932
hCLASP2          DKFGSENVKMIQDSGKVNPKDLDSKYAYIQVTHVIEFFDEKELQERKTEFERSHNIRRFM 1770
hCLASP7          ERFGDDVVEIIKDSYPVDKSKLDSQKAYIQITYVEEYFDTYELKDRVTYFDRNYGLRTFL 1851
hCLASP1          DKFGADNVKIIQDSNKVNPKDLDPKYAYIQVTYVTEFFEEKEIEDRKTDFEMHHNINRFV 1972
                 :  ** : *::*:**  *:  .**.:  *:**:*.* *:*:  *: :* * *:  ..:  *:
                                                 ITAM                  DOCK motif
hCLASP4          FEAPYTLSGKKQGCIEEQCKRRTILTTSNSFPYVKKRIPINCEQQINLKPIDGATDEIKD 1859
hCLASP5          YTTPFTLEGRPRGELHEQYRRNTVLTTMHAFPYIKTRISVIQKEEFVLTPIEVAIEDMKK 1870
hCLASP3          YCTPFTLDGRAHGELHEQFKRKTILTTSHAFPYIKTRVNVTHKEEIILTPIEVAIEDMQK 1992
hCLASP2          FEMPFTQTGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQHHTDLNPIEVAIDEMSK 1830
hCLASP7          FCTPFTPDGRAHGELPEQHKRKTLLSTDHAFPYIKTRIRVCHREETVLTPVEVAIEDMQK 1911
hCLASP1          FETPFTLSGKKHGGVAEQCKRRTILTTSHLFPYVKKRIQVISQSSTELNPIEVAIDEMSR 2032
                 :  *:*  *: :* : ** :*.*:*:: : **:*:*: :  .    *.*:: * :::.
                          Coiled-coil
hCLASP4          KTAELQKLCSSTDVDMIQLQLKLQGWVSVQVNAGPLAYARAFLNDSQASKYPPKKVSELK 1919
hCLASP5          KTLQLAVAINQEPPDAKMLQMVLQGSVGATVNQGPLEVAQVFLAEIPADPKLYRHHNKLR 1930
hCLASP3          KTQELAFATHQDPADPKMLQMVLQGSVGTTVNQGPLEVAQVFLSEIPSDPKLFRHHNKLR 2052
hCLASP2          KVAELRQLCSSAEVDMIKLQLKLQGSVSVQVNAGPLAYARAFLDDTNTKRYPDNKVKLLK 1890
hCLASP7          KTRELAFATEQDPPDAKMLQMVLQGSVGPTVNQGPLEVAQVFLAEIPEDPKLFRHHNKLR 1971
hCLASP1          KVSELNQLCTMEEVDMISLQLKLQGSVSVKVNAGPMAYARAFLEETNAKKYPDNQVKLLK 2092
                 *. :*          *   **: *** *.  ** **:  *:.** :   .    .: . *:
                                    Coiled-coil
hCLASP4          DMFRKFIQACSIALELNERLIKEDQVEYHEGLKSNFRDMVKELSDIIHEQILQEDTMHSP 1979
hCLASP5          LCFKEFIMRCGEAVEKNKRLITADQREYQQELKKNYNKLKENLRPMIERKIPELYKPIFR 1990
hCLASP3          LCFKDFTKRCEDALRKNKSLIGPVQKEYQRELGKLSSP---------------------- 2090
hCLASP2          EVFRQFVEACGQALAVNERLIKEDQLEYQEEMKANYREMAKELSEIMHEQICPLEEKTS- 1949
hCLASP7          LCFKDFCKKCEDALRKNKALIGPDQKEYHRELERNYCRLREALQPLLTQRLPQLMAPTP- 2030
hCLASP1          EIFRQFADACGQALDVNERLIKEDQLEYQEELRSHYKDMLSELSTVMNEQITGRDDLSKR 2152
                   *:.*    *  *:  *: **    * **:. :
                                    PDZ ligand
hCLASP4          WMSNTLHVFCAISGTSSDRGYGSPHYAEV-- 2008
hCLASP5          VESQKRDSFHRSSFRKCETQLSQGS------ 2015
hCLASP3          ------------------------------
hCLASP2          VLPNSLHIFNAISGTPTSTMVHGMTSSSSVV 1980
hCLASP7          --PGLRNSLNRASFRKADL------------ 2047
hCLASP1          GVDQTCTRVISKATPALPTVSISSSAEV-- 2180
```

FIG. 13
6 of 6

Human CLASP-2 expression in T cells upon activation

A)



B)

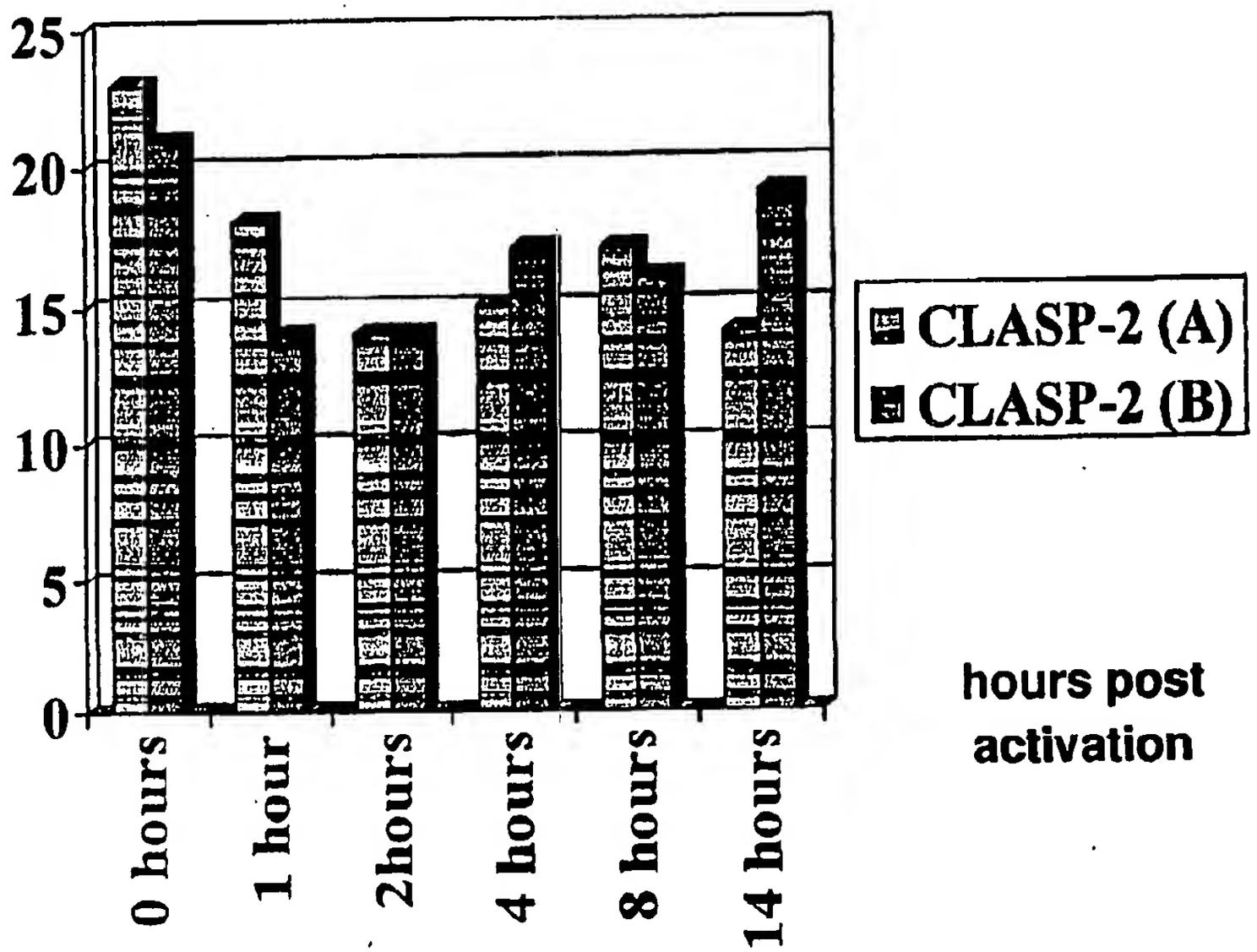


FIG. 14